Further Characterisation of the University of Canterbury Paediatric Auditory-Visual Matrix Sentence Test

A thesis submitted in partial fulfilment of the requirements for the Degree of

Master of Audiology

in the School of Psychology, Speech and Hearing

at the University of Canterbury

H. R. Taylor

University of Canterbury

2021

Acknowledgements

Thank you to all of the lecturers and clinical staff belonging to the Audiology department for your guidance and your patience over the past two years. I am grateful to have been supported by such a knowledgeable, supportive, and encouraging team.

In particular, I would like to express a big thank you to my primary supervisor Professor Greg O'Beirne. This research could not have been completed without all of the long hours that you have spent developing the UCAMST and the UCAMST-P. Thank you for the ongoing guidance, motivation, and enthusiasm; I am hugely grateful.

Thank you also to my secondary supervisor, Associate Professor Rebecca Kelly-Campbell. Your statistical expertise has been invaluable and I appreciate all the time you spent helping me to work through my statistics.

I would like to acknowledge each of the participants who have volunteered their time at some point along the ongoing research into the UCAMST and UCAMST-P. Without your commitment, none of this research would have been possible.

Lastly, I would like to express my sincerest appreciation to all of my classmates. I could not have hoped for a better group to work with over the past two years. I have learnt so much from you all and I look forward to seeing where we all end up.

Abstract

The impact of a hearing impairment is significant. It can impede an individual's ability to hear and understand speech at normal conversational levels, which is considered an integral part of daily communication. A standard audiological assessment is employed to determine an individual's hearing ability, which includes a test of speech recognition. However, such tests frequently present speech as single words in a quiet clinical setting. This artificial listening situation restricts the application of the test as it does not accurately represent the individual's ability in real-life listening environments. The University of Canterbury Auditory-Visual Matrix Sentence Test was developed in New Zealand English by a team led by Professor Greg O'Beirne, to fill the need for a New Zealand speech test that more accurately determines an individual's true ability to understand verbal communication in their everyday life. Due to the cognitive demands of typical matrix sentence tests, a paediatric version (UCAMST-P) was produced by Jenkins-Foreman (2018). The current study sought to reanalyse corrected data from Jenkins-Foreman (2018) pertaining to equivalence of UCAMST-P sentence lists and conditions in the auditory-alone modalities. Following this, the present study aimed to assess whether the UCAMST-P might be more suitable for populations with a reduced working memory capacity through the use of response time and response order data in normal-hearing participants, previously obtained by Jenkins-Foreman (2018). Evaluation of the UCAMST-P revealed that sentence lists within the AA open- and closed- set conditions, are indeed equivalent. As expected, the auditory-alone speech reception thresholds (SRTs) obtained in the open- and closed- set conditions were not equivalent to each other. Recalculation of Lay's (2019) results using the new data revealed similar results for list and condition equivalency, within and

between each condition. Testing the UCAMST under low SNR conditions gave rise to rapid responses and poor accuracy, providing evidence of rapid-guessing behaviour. Button presses for both tests were more sequential under high SNR conditions, and practice appears to have had little effect on either test.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AA	Auditory-Alone
ANOVA	Analysis of Variance
AV	Audio-Visual
BKB-SIN	Bamford-Kowal-Bench Speech-in-Noise
CVC	Consonant-Vowel-Consonant
dB	Decibels
dB A	A-Weighted Decibels
dB HL	Decibels in Hearing Level
dB SNR	Decibels Signal-to-Noise Ratio
dB SPL	Decibels Sound Pressure Level
ELU	Ease of Language Understanding
FINSIMAT	Finnish Simplified Matrix Test
FMST	Finnish Matrix Sentence Test
HI	Hearing Impairment
HINT	Hearing in Noise Test
HINT-C	Hearing in Noise Test for Children
Hz	Hertz
IRT	Inter-Response Time
ITAMatrix Test	Italian Matrix Test
KTT	Kendall Toy Test
MST	Matrix Sentence Test
NH	Normal Hearing

NZ	New Zealand
NZAS	New Zealand Audiological Society
NZHINT	New Zealand Hearing in Noise Test
OlKiSa	Oldenburger Kinder-Satztest
OLSa	Oldenburg Satztest
OLSAf	Oldenburger Satztest - Female
PI	Performance-Intensity
PPMST	Polish Paediatric Matrix Sentence Test
PSI	Paediatric Speech Intelligibility
PTA	Pure-tone Audiometry
QuickSIN	Quick Speech-in-Noise
SilMax	Simplified Italian Matrix Sentence Test
SLM	Sound Level Meter
SNR	Signal-to-noise ratio
SPSS	Statistical Package for the Social Sciences
SRT	Speech Recognition Threshold
UCAMST	University of Canterbury Auditory-Visual Matrix Sentence Test
UCAMST-P	University of Canterbury Auditory-Visual Matrix Sentence Test – Paediatric
UNHSEIP	New Zealand Universal Neonatal Hearing Screening and Early Intervention Programme
VA	Visual-alone
WMC	Working Memory Capacity

Chapter One

1.1 Background

Hearing allows engagement in verbal conversations, appreciation for music, and the ability to process and selectively listen amongst the noise of the surrounding world; all while ensuring safety and signalling potential danger. However, hearing impairment (HI) is a common and widespread disability in society (Olusanya et al., 2014). The total proportion of New Zealanders with a HI is unclear. Statistics NZ (2013) estimated that 9% of New Zealander's had a HI, while Stevens and colleagues (2011) estimated that 19% would have a HI by 2016. For those affected by a HI, a leading complaint is reduced ability to hear and understand speech, particularly with simultaneous background noise (Healy & Yoho, 2016). These individuals must increase their listening effort to follow conversation which can lead to loss of confidence and in some cases social withdrawal (Alhanbali et al., 2017; Beechey et al., 2020).

Children with a HI are particularly susceptible to developmental concerns as they can miss certain speech sounds. This can impede the child's development of oral language proficiency, adversely affecting their communication skills and leading to poorer social participation and interaction (Hadjikakou et al., 2008; Jalilevand et al., 2018; Lang-Roth, 2014; Patel et al., 2014; Sarant et al., 2008). A systematic review and meta-analysis by Roland and colleagues (2016) revealed that children with a HI had clear reductions in two quality of life domains: school activities and school interactions, both of which are paramount for ongoing social development and learning. It is unsurprising that these children have a greater risk of developing emotional and behavioural difficulties compared with their normal hearing counterparts (Stevenson et al., 2015).

Pure-tone audiometry (PTA) is typically used to diagnose a HI. However, HI affects individuals differently and PTA does not explicitly identify how an individual performs in an everyday environment with regards to their speech recognition capacity. Speech recognition testing, which employs speech stimuli instead of pure tones, provides greater insight into one's ability to detect, recognise, and understand speech (Mendel, 2008). The outcomes of such speech tests are essential to initiate a personalised and effective management and rehabilitation plan.

Owing to shortfalls in current New Zealand speech recognition tests, the UCAMST (O'Beirne et al., 2015; Trounson, 2012) and the UCAMST-P (Jenkins-Foreman, 2018) have been produced to accurately assess an individual's day-to-day ability to understand sentences in background noise – a situation that is similar to many listening environments. However, due to the cognitive demands of sentence testing (Cervera et al., 2009; Daneman & Carpenter, 1980; Theunissen et al., 2009), research is necessary to build the evidence base supporting the use of the UCAMST-P rather than the UCAMST for populations with a developing, or reduced working memory capacity (WMC).

1.2 Audiological Assessment in New Zealand

Correctly identifying the presence, source, and degree of a HI permits a more informed approach to initiating management strategies and ongoing rehabilitative care (Madell & Flexer, 2014; Neumann et al., 2012). A test battery approach incorporates several different tests and is employed to target the integrity of each distinct segment of the auditory system. This provides differential information about the nature and source of an auditory deficit and permits cross-checking of results for improved validity (Norrix, 2015). The New Zealand Audiological Society (NZAS; 2016) suggests that audiological test batteries should consist of both subjective testing (pure-tone audiometry (PTA) and speech recognition testing) and objective procedures (tympanometry, acoustic reflexes, and otoacoustic emissions).

1.2.1 Pure-tone Audiometry

PTA is currently regarded as the gold standard of hearing assessment and is routinely performed to determine an individual's hearing thresholds across the frequencies necessary for hearing in daily life (Fredriksson et al., 2016). It is a behavioural test requiring a listener to press a button in response to hearing pure tones. In New Zealand (NZ), a modified Hughson-Westlake down-up procedure is employed to systematically vary the intensity of each pure tone stimulus and locate the threshold; defined as the lowest intensity level at which a response occurs in at least half of the ascending trials (Hughson & Westlake, 1944; NZAS, 2020). Typically, the thresholds are plotted on an audiogram mapping the individual's hearing sensitivity as a function of frequency (Hz) which is used to guide a clinicians decision on management and rehabilitation strategies (NZAS, 2020; Schlauch & Nelson, 2015). While PTA is essential for diagnosis, it does not fully predict listener's real-world communication ability or the likelihood of success with rehabilitation, therefore, additional testing procedures add further diagnostic value.

1.2.2 Speech Audiometry

Speech recognition tests are conducted using speech stimuli to more accurately assess an individual's communication capacity (Konkle & Rintelmann, 1983; McArdle & Hnath-Chisolm, 2015). They provide valuable supplementary information to the PTA; acting as a cross-check of the PTA reliability and adding further diagnostic information to the nature of the HI (Jerger & Hayes, 1977; McArdle & Hnath-Chisolm, 2015). Speech recognition tests have pivotal impacts on subsequent rehabilitation or management options, for example, hearing aids or cochlear implant candidacy, need for assistive listening devices, and the diagnoses of pathologies, such as auditory neuropathy and auditory processing disorders (Hoppe et al., 2015; Iliadou, et al., 2017; Zeng & Liu, 2006).

1.2.2.1 CVC Words Lists

NZ has adopted the meaningful Consonant Vowel Consonant (CVC) words lists to conduct adult speech audiometry in accordance with the New Zealand Audiological Society (NZAS) best practice guidelines (Boothroyd, 1968; NZAS, 2016). The test contains ten lists of ten monosyllabic and phonetically balanced words. Each word in a list is presented at the same intensity, and each word is presented following the carrier phase: “say ___” (e.g. “say pass”). The individual must verbally repeat each word, or part of the word, as they hear it and scores are recorded as the total number of correctly identified phonemes per list. The score (%) for each list is plotted against its presentation level (dB HL) and a sigmoidal line of best fit is drawn between the points – a plot referred to as the performance/intensity (PI) function. From this function, the speech recognition threshold (SRT; the speech intensity level correlating to 50 percent intelligibility) can be derived and used to cross-check the pure tone thresholds (Boothroyd, 2008; Brand & Kollmeier, 2002; Mendel, 2008). As the PI function illustrates an individual’s hearing ability at various intensities, it is helpful for guiding rehabilitative decisions. For example, those with retro-cochlear pathologies or cochlear dead regions typically have the poorest speech recognition and may not be able to achieve greater than 90% discrimination, regardless of intensity (McArdle & Hnath-Chisolm, 2015). In such cases, amplification alone will not significantly aid in speech recognition and implementing additional auditory training and communication strategies should be advised (Benson et al., 2018).

However, the CVC words lists and various other speech recognition tests have some limitations pertaining to their resemblance to everyday communication. The reliability of the

CVC words lists, for example, depends on controlled testing environments with low levels of ambient noise to avoid artificially increasing hearing thresholds (Boothroyd, 1968; Maclennan-Smith et al., 2012; NZAS, 2016). Consequently, using solely this speech-in-quiet test approach will make a clinician less equipped to predict their client's success with amplification in noisy, everyday environments (Beattie et al., 1997). Additionally, the simplicity of the speech in quiet test means that it has varying ability to differentiate between those with normal hearing (NH) from a those with a mild HI (Beattie et al., 1997). Not being able to differentiate between those with and without a HI is detrimental as many adults with a mild HI can still benefit from amplification (Kelly-Campbell et al., 2014). On the other hand, HAs can augment the difficulties experienced in background noise by some individuals with a HI (Kelly-Campbell & Lessoway, 2015). In both cases, implementing speech-in-noise testing can improve rehabilitative outcomes (Beattie et al., 1997).

1.3 Speech in Noise Testing

For those with a HI, understanding speech in noise is complicated due to several physiological factors including: elevated hearing thresholds, loss of frequency selectivity, the presence of loudness recruitment, and impaired temporal resolution (Legris et al., 2018; Peters & Moore, 1992). Understanding speech in noise, therefore, is a primary concern of those with a HI (Beattie et al., 1997; Billings & Madsen, 2018; Dirks et al., 1982; Healy & Yoho, 2016). To test an individual's capacity to communicate in every-day, noisy environments, some speech recognition tests can be conducted in the presence of noise (Hagerman, 1982; Portnuff & Bell, 2019; Trounson, 2012).

1.3.1 Procedure

Speech in noise tests involve the listener repeating various words or sentences in simultaneous background noise while the clinician alters the intensity of the signal or the noise (Legris et al., 2018). This change in intensity alters the signal to noise ratio (SNR), defined as the difference between the intensity of the speech and the noise (dB SNR) (Legris et al., 2018). The purpose of such tests is to locate the speech perception threshold (the SNR where the listener correctly identifies half of the test stimuli), and the maximum speech perception threshold (the SNR where the listener correctly identifies all the stimuli in a given list) (Legris et al., 2018). This information guides rehabilitative decision-making and can be used to counsel the client on different approaches (Beattie et al., 1997; Wilson et al., 2007). While such tests exist for clinical use, their uptake is slow and remains in beginning stages (Wagener & Brand, 2005).

1.3.2 Acoustic Masking Noise

To simulate everyday communication environments in a clinic, masking noise is necessary. Multi-talker babble and continuous speech-shaped noise are two methods used with speech testing measures (Killion et al., 2004). Speech-shaped noise has similar spectral content to the target signal which can lend greater sensitivity than that provided by fluctuating babble noise (Francart et al., 2010; Wagener & Brand, 2005). Speech-shaped noise is therefore desired in research settings when discriminating between two variables (Nilsson et al., 1994; Plomp & Mimpen, 1979a; Wagener & Brand, 2005). On the other hand, multi-talker babble mimics background voices, which tends to be the most noticeable type of everyday background noise. This makes it more suited to clinical speech testing and achieves greater face-validity than speech-shaped noise (Francart et al., 2010, Hopkins & Moore, 2009; Killion, 2002, Killion et al., 2004).

1.3.3 SRT and Psychometric Functions

In tests that incorporate background noise, a psychometric function illustrates the listener's performance as a percentage intelligibility score (%), as a function of the SNR. These psychometric functions are sigmoidal (s-shaped) and are typically described by the threshold level and the slope (Figure 1). Similar to the PI function produced in speech-in-quiet testing, the SRT for open-set tests is located at the point on the psychometric function of 50 percent intelligibility.

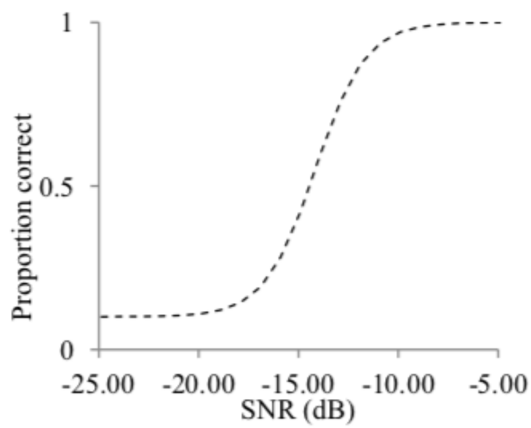


Figure 1. Sigmoidal shape of a psychometric function measuring the proportion of correct responses as a function of the SNR (dB). *Image from McClelland (2015, p.12).*

In noise, the accuracy of the SRT is dependent on the slope of the psychometric function at the location of the SRT. A strict inverse relationship exists between the psychometric function slope, and the accuracy of the SRT; a steeper slope at the SRT results in a lower standard deviation of the SRT (Ozimek et al., 2012). The sensitivity of the test is regulated by the slope of the psychometric function; the steeper the function, the greater the sensitivity (Figure 2). A higher sensitivity test (steeper psychometric slope) is preferred as fewer trials are necessary to accurately locate the SRT (Francart et al., 2010).

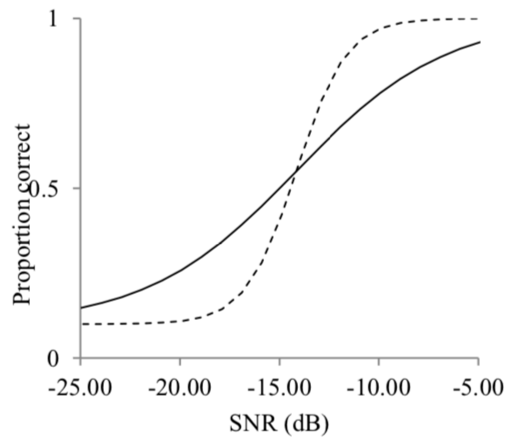


Figure 2. Comparison of psychometric functions with steep (dashed line) and shallow (solid line) slopes. *Image from McClelland (2015, p.13).*

A test with higher sensitivity also means that a slight change in stimulus level causes a considerable change in the measured intelligibility value. This can be used to illustrate the potential benefit of slightly adjusting the SNR through amplification and can guide a clinician's discussion with a client regarding the realistic expected benefits of varying amplification methods (Brand & Kollmeier, 2002; Wilson et al., 2007).

1.3.4 Tracking the SNR

Two different methods are used to estimate the SRT in speech perception tests: non-adaptive and adaptive measures. The non-adaptive method uses varying levels of intensity that are calculated before the assessment begins, and do not change during the test. In an adaptive procedure the stimulus level for each trial is changed based on the previous response (Levitt, 1971). A common adaptive procedure to measure thresholds is the up-down (staircase) response (Brown, 1996; Levitt, 1971; Plomp & Mimpen, 1979) where the stimulus level is adjusted up or down by a consistent value depending on the listener's preceding response (Plomp & Mimpen, 1979). A correct response, for example, might result in the presentation level being reduced by 2 dB, increasing the difficulty for the next presentation, and vice-versa. These adaptive tracking methods are reliable and efficient regardless of any changes in signal or noise and increase the overall reliability and efficiency of the test while also avoiding floor and ceiling effects (Lay, 2019; Wagener & Brand, 2005).

1.4 Sentence Testing

1.4.1 Single-word or Sentence Stimuli

Speech recognition tests typically require individuals to listen for single phonemes, words, or sentences (Dietz et al., 2014). Listeners are then scored on their ability to repeat, or otherwise identify, the presented stimuli. This is used to determine their speech perception ability. Shorter stimuli like phonemes or syllables are occasionally preferred for paediatric populations as the test picks up on repetitive phoneme errors and illustrates how a child discerns specific phonological variations (Neumann et al., 2012). Similarly, shorter stimuli can pick up word recognition errors due to confusion of consonants, such as /s/, /f/, or /th/, which are associated with a high-frequency, age-related HI (Stelmachowicz et al., 2004).

Conversely*,* sentence-stimuli are beneficial as they mimic a normal conversation. Compared with single-words, sentence stimuli provides a broader dynamic range attributed to the varying fluctuations, intonations, pauses, and temporal and contextual cues present in real-world conversational speech, which increases the test's validity (Dietz et al., 2014; Killion et al., 2004; Nilsson et al., 1994). Sentences also afford more words per trial than single-word stimuli which can increase the test efficiency (Hagerman, 1976; Hochmuth et al., 2012). Hagerman (1976) illustrates that doubling the number of words per sentence increases the accuracy of the speech perception test by √2 (1.41). As a result, sentence-stimuli produce steeper psychometric functions than with single-word stimuli, leading to more accurate SRT measurements (discussed in section 1.3.3) (Hagerman, 1976; McArdle et al., 2005).

The most common sentence tests are Plomp-type sentence tests which consist of phonemically balanced sentence lists based off everyday conversation, yet carry no consistent grammatical structure (Plomp & Mimpen, 1970; Wong et al., 2019). Neighbouring words within sentence-stimuli, however, provide some contextual cues (semantic, syntactic, and or prosodic) to the listener so that words can be extrapolated even if they were not heard correctly (Hutcherson et al., 1979). Neighbouring words reduce the accuracy of the speech recognition measure as the listener becomes less reliant on the exact acoustic properties of the signal and more reliant on their cognition (Kalikow et al., 1977; Madell & Flexer, 2014; Wilson et al., 2007). Sentence stimuli is also more taxing on individual's working memory capacity (WMC) as the listener must hold onto the each word until the end of the sentence before recalling each word (Cervera et al., 2009; Daneman & Carpenter, 1980; Theunissen et al., 2009). Nonetheless, current literature suggests that using sentence-stimuli for speech recognition testing more accurately estimates how the listener performs in daily life, primarily due to its resemblance of every-day communication (Hochmuth et al., 2012; Killion et al., 2004; Ozimek et al., 2009; Theunissen et al. 2009).

1.4.2 Matrix Sentence Tests

A Matrix Sentence Test (MST) is an alternative sentence-based speech recognition measure which uses fixed sets of syntactically-correct sentences that are semantically unpredictable for the listener (Hagerman, 1982; Hochmuth et al., 2012; Meister, 2016). The five-word sentences (consisting of a name, verb, number, adjective, and object) are derived from a five by ten-word matrix of words, giving rise to an essentially unlimited bank of up to 10^5 (100,000) possible sentences. In modern implementations of this test (after Wagener et al., 1999a-c), the words of the matrix are recorded separately so that there is co-articulation between words.

Current literature agrees that MSTs assess a combination of signal processing, background noise, and individual WMC more consistently than with hearing in noise (HINT) testing (Plomp-type sentences) (Rudner et al., 2009; Rudner et al., 2011). These findings are parallel to those of other literature which indicate that linguistic factors, such as context, mitigate the influence of cognitive factors on the outcome of the speech recognition test (Wingfield et al., 2015). Therefore, matrix sentence-style testing is likely more sensitive regarding cognitive affects, such as WMC compared with Plomp-type sentences.

1.5 The UCAMST: A New Zealand Matrix Sentence Test

Originally produced in Swedish by Hagerman (1982), various MSTs have since been produced internationally owing to performances on the test being impacted by the speakers accent and pronunciation (Hochmuth et al., 2012). MSTs are currently available in: Danish (Wagener et al., 2003), Dutch (Houben et al., 2014), Finnish (Dietz et al., 2014), French (Jansen et al., 2012), German (Wagener et al., 1999), Italian (Puglisi et al., 2015), Mandarin (Tao et al., 2017), Russian (Boboshko et al., 2013), Polish (Ozimek et al., 2010), Spanish (Hochmuth et al., 2012), and Turkish (Zokoll et al., 2015). However, New Zealand English is different to other forms of English largely due to differences in formant structure and vowel pronunciation (Maclagan & Hay, 2007).

As a result, The University of Canterbury Matrix Sentence Test (UCAMST) was established in New Zealand English by O'Beirne and Trounson for clinical purposes in NZ (O'Beirne et al., 2015; Trounson, 2012). It was adapted from the British English MST (Hall, 2006) and uses recorded sentence stimuli spoken by a native NZ English speaker, with one of two different background noises: constant speech-shaped noise (King, 2010; Stone, 2016), or six-talker babble (Ripberger, 2018; Spencer, 2011). The five-word sentences are constructed from the matrix produced by Trounson (2012; Figure 3).

Name	**Verb**	**Quantity**	**Adjective**	**Object**
Amy	bought	two	big	bikes
David	gives	three	cheap	books
Hannah	got	four	dark	coats
Kathy	has	six	good	hats
Oscar	kept	eight	green	mugs
Peter	likes	nine	large	ships
Rachel	sees	ten	new	shirts
Sophie	sold	twelve	old	shoes
Thomas	wants	some	red	spoons
William	wins	those	small	toys

Figure 3. The UCAMST word matrix (Trounson, 2012).

The UCAMST can be administered under either a closed-set response format where the individual responds by selecting what they heard from a visible matrix, or an open-set response format where the response is given verbally without visual clues. The test has previously been cross validated with the CVC words list, and a significant relationship was found between the

SRTs of the two tests suggesting potential capacity for the UCAMST to replace the current CVC word recognition test in NZ clinical practice (Ripberger, 2018).

1.5.1 Tracking Procedure

The alternative adaptive tracking procedure employed for the UCAMST was created by Brand and Kollmeier (2002) to estimate the slope of the psychometric function and the SRT concurrently. The method works by tracking two points on the psychometric function, described as the "pair of compromise", and generally refers to the 20 and 80 percent correct points. Measuring the listener's SRT speeds the process of locating the threshold and this efficiency makes it preferential to non-adaptive procedures, whilst also avoiding floor and ceiling affects that can occur when scores are based on percentage correct (Levitt, 1978).

1.5.2 Auditory-Visual Component

The New Zealand English and Malay versions of the UCAMST are unique to other international matrix tests as they have the option to feature a visual recording which displays the speaker's face as they present the sentence stimuli. Granting access to both auditory and visual speech signals gives complementary information which is combined so that speech understanding under auditory-visual (AV) conditions can exceed that of speech understanding in auditory-alone (AA) conditions, especially in situations with reduced SNRs (McClelland, 2015; Sommers et al., 2005; Sumby & Pollack, 1954; Tye-Murray, et al., 2007). Although, Llorach and colleagues (2019) created an AV version of the German MST (AV-OLSAf) which showed that in noise, the AV benefit was only 5 dB over the AA condition. Additionally, the AV modality suffered from floor and ceiling effects which were closely related to the listener's speech reading abilities (Llorach et al., 2019). Regardless, the ability to assess an individual's performance under both conditions can give the test greater application, as it provides comprehensive insight into how the individual performs in everyday listening environments which, in turn, can assist in producing a tailored rehabilitation program (Tye-Murray et al., 2007). These recordings have been extensively examined by Trounson (2012) and McClelland (2015) to ensure that the sentences appeared natural following their assembly from a base matrix of individual words.

1.5.3 Sentence List Equivalence and Normalisation

Assessing the sentence list equivalency and normalising test material so that each word in the matrix has equal intelligibility in noise has been a large focus in research pertaining to the UCAMST development (Jenkins-Foreman, 2018; McClelland, 2015; Ripberger, 2018; Stone, 2016). Regardless of response format, under the AV modality the UCAMST sentence lists are equivalent with respect to SRT and slope, and therefore the UCAMST sentence list stimuli can be used interchangeably between the open and closed set formats (Jenkins-Foreman, 2018; Stone, 2016). Conversely, under the AA modality some lists were significantly different with respect to SRT and slope (Jenkins-Foreman, 2018). Significant differences were observed in the SRT between the open- and closed-set response formats in both presentation modes.

Regardless of sentence stimuli affording a more accurate representation of daily listening environments, the listener must listen and retain each word for the duration of the sentence in their auditory memory before responding, placing more demands on the listener's cognitive capacity (Theunissen et al., 2009). It is well documented that WMC is reduced in paediatric, and senior adult populations (Foo et al., 2007; Hällgren et al., 2001; Humes et al., 2012; Jerger et al., 1991), and this must be considered when implementing a sentence-test as not to confound an individual's ability to hear with ability to remember.

1.6 Cognition: WMC and Attention

1.6.1 WMC

While several models exist to explain WMC, all conclude that it is a limited capacity cognitive system which involves active manipulation of information that is already being held in focal attention (Alloway et al., 2006; Baddeley & Hitch, 1974; Cowan, 2008; Glisky, 2007; Jonides et al., 2008; Park & Hedden, 2001; Reuter-Lorenz & Sylvester, 2005). The most common model of WMC, created by Baddeley and Hitch (1974), breaks WMC into three components: the phonological loop (auditory information), the visuo-spatial sketchpad (visual information), and the central executive (supervision of the other components). Of importance for audiological purposes is the phonological loop which is activated by auditory stimuli and may be referred to as auditory memory (Baddeley & Hitch, 1974). The phonological loop itself works by recalling speech sounds in their temporal order while maintaining and regenerating traces of memory by repeating the words in a loop (Baddeley, 1986; Baddeley & Hitch, 1974). This is different to short-term memory, which involves simply maintaining information for a short time period (Cowan, 2008; Glisky, 2007; Jonides et al., 2008). Importantly, WMC is typically reduced in young children school-aged children (Gathercole et al., 2004; Lendinez et al., 2015; Luna et al., 2004) and with increasing age in late adulthood (Glisky, 2007; Humes et al., 2012; Jerger et al., 1991; Pliatsikas et al., 2018).

1.6.2 WMC and Sentence Testing

As an auditory signal and other background noise is received by the auditory system, implicit sensory (bottom-up processes), are employed to describe the perception of the auditory signals (Pichora-Fuller et al., 1995). The listener then uses higher-order (top-down) cognitive processes including WMC and attention (Beechey et al., 2020; Rönnberg et al., 2013), to optimize their understanding of the signal with reference to prior knowledge (Conway et al., 2009).

When conducting speech recognition testing there is an increased cognitive load, particularly with background noise or when the auditory signal is compromised (Akeroyd, 2008; Souza et al., 2015). The Ease of Language Understanding Model (ELU; Rönnberg et al., 2013) explains this phenomenon. For a listener with NH, when a signal is presented with no masking noise the perception and understanding of speech occurs automatically and effortlessly through an implicit process. However, if the speech input is compromised by way of a poor signal, HI, or background masking noise, an explicit process that exhausts cognitive resources is employed to make use of the compromised signal. The ELU model proposes that WMC may be utilised to restore the elements of speech input and improve speech perception in adverse listening environments with a low SNR (Rönnberg et al., 2013). Notably, in taxing cognitive resources with speech processing, greater listening effort is required (McGarrigle et al., 2014).

For sentence-based speech tests, particularly those in noise, auditory speech processing and working memory are critical as the listener must process, unravel, and store the signal information for the duration of the sentence before identifying the constituent words (Cervera et al., 2009; Daneman & Carpenter, 1980). As WMC is in the order of seconds (provided rehearsal and active maintenance are avoided), a longer sentence has a longer duration, yielding greater taxes on WMC. Meister (2016) found a stronger relationship between cognitive functions, including WMC, when using sentence-based material and or background masking noise, compared with single-word tests in quiet. The effect of WMC on speech recognition tasks has been researched by van Rooij and Plomp (1990) and Akeroyd (2008) with both studies revealing clear associations between weakened WMC and increased SRTs. Cognition may then confound a listener's ability to understand complex stimuli in challenging listening environments. Consequently, WMC must be considered when conducting speech recognition testing, particularly when the test necessitates recall of full sentences which may be taxing on children and elderly populations.

1.6.3 Testing WMC

A handful of tests exist to capture individual differences in peoples WMC. The most well-known and common test of measuring WMC is the complex span test, where a researcher presents lists of items (digits or words) of increasing length (Conway et al., 2005; Daneman & Carpenter, 1980). An individual's WMC is determined by the longest list length that they can remember accurately, on at least half of the trials. Miller (1956) proposed that humans have a memory span of roughly seven items ± two, however, this has been widely contested and shown to differ between populations and stimuli. Some literature suggests that memory span can change based on the category of stimuli used: a digit-span of seven, letter-span of six, and word-span of five (Crannell & Parish, 1957), while others estimate that WMC is even less; approximately four items in young adults and fewer in children and older adults (Cowan, 2001). It is also unclear how the effect of degraded listening conditions, like a reduced SNR, affects word-recall. Nonetheless, if approximations of word-recall length have been made between four to five under non-degraded listening conditions, the more cognitively demanding five-word UCAMST may be at risk of being influenced by WMC, particularly for younger and older populations. It would be beneficial, therefore, to administer a test of WMC such as the digit span test or letter number sequencing (Calamia et al., 2012), prior to conducting testing pertaining to the effects of WMC on the UCAMST and UCAMST-P, in order to establish each participant's baseline WMC to accurately discern the differences in cognitive loads between the two tests.

1.6.4 Working Memory in Older Populations

For tasks requiring active manipulation, reorganisation, or integration of information held in their working memory, older adults experience significant deficits (Glisky, 2007). While research on older adults typically refers to adults aged over 60 years old (Cansino et al., 2020; Cansino et al., 2013; Klencklen et al., 2017), WMC likely starts to decline slowly from the age of thirty (Cansino et al., 2020; McNab et al., 2015). Several theories seek to suggest why this occurs, including: a reduction in processing and attentional resources (Craik, 1986; Craik & Byrd, 1982), a general slowing of processing information (Salthouse, 1995), and poor inhibitory control preventing deletion of information in working memory to make room for new information (Hasher & Zacks, 1988, Hasher et al., 1999; May et al., 1999; Park, 2000). Although the underlying cause of WMC reductions remains contested, auditory processing clearly declines with age and may be attributed to various cognitive, peripheral, and central aspects (Humes et al., 2012; Jerger et al., 1991).

For individuals with a HI, WMC can be reduced even further. Lin et al. (2013) found a significant relationship between a person's HI and their cognitive function: the greater the hearing impairment, the greater the risk of cognitive deterioration. This supports earlier literature which drew clear associations between increasing age and declines in WMC (Foo et al., 2007; Hällgren et al., 2001). Additional research indicates that a reduction in memory capacity can increase those individuals' SRTs (Theunissen et al., 2009; van Rooij & Plomp, 1990). As adults over 65 years of age already have a greater prevalence of memory and hearing impairments (Li et al., 2006) and tests requiring WMC are by nature, cognitively demanding (Glisky, 2007), there is a real need to employ measures of speech recognition that locate the SRT accurately and efficiently in older populations (Newman & Sandridge, 2004).

1.6.5 Working Memory in Children

Auditory memory capacity is also affected in childhood and adolescence as working memory develops in this time (Lendinez et al., 2015; Luna et al., 2004). The efficiency with which information can be updated and amended is progressively developed from birth throughout childhood as working memory improves, and as such, children's performance on auditory listening tasks continually improves as a child matures (Gathercole et al., 2004). Due to greater developmental variation in younger age groups, there is also increased variability in the speech recognition scores of younger children compared to older children (Holder et al., 2016; Ng et al., 2011; Wilson et al., 2010). As with senior populations, these reductions in memory capacity can increase speech recognition thresholds, raising the importance of auditory memory when implementing speech recognition tests (Cervera et al., 2009; McArdle et al., 2005; Theunissen et al., 2009; van Rooij & Plomp, 1990; Wilson et al., 2007).

Children with a HI are consistently shown to have delays in their auditory memory development (Dawson et al., 2002; Pisoni & Cleary, 2003; Pisoni et al., 2008). There is some agreement across the literature that a 'sensitive' period of auditory memory development occurs at some point before age eight (King et al., 2002; Purdy et al., 2002; Tremblay, et al., 1997). Therefore, children who experience auditory deprivation due to a HI during early developmental years may have impaired ability to optimise their WMC and, as such, a HI may appear augmented when implementing speech recognition testing that is taxing on cognitive resources (Davidson et al., 2019). A concise speech recognition test that reflects an individual's true ability to listen in noisy environments is therefore warranted (Cervera et al., 2009; McArdle et al., 2005; van Rooij & Plomp, 1990; Wilson et al., 2007).

1.7 Paediatric Speech Recognition Testing

The current test battery for paediatric populations is near-identical to that of adults, however. Behavioural tests, including PTA and speech audiometry, must be adapted to suit the child's age, physical ability, cognitive function, and presence of any other confounding disabilities (Diefendorf & Wynne, 2004; NZAS, 2015). Children can have varying levels of language development, particularly those with a HI, who are susceptible to delayed language development and a smaller vocabulary (Cupples et al., 2017; Percy-Smith et al., 2018; Walker et al., 2015). Therefore, speech recognition tests must not be swayed by a child's vocabulary comprehension, higher-level language abilities, speech pronunciation skills or phonological knowledge (Kosky & Boothroyd, 2003; Neumann et al., 2012). As the length of the test is primarily dictated by the child's fatigue, the test must also account for the child's attention levels, how interesting the test is for the child, and the efficiency of the test (Neuman et al., 2012). Therefore, currently available speech intelligibility tests designed for adult use are not always desirable for paediatric assessment (Wagener & Kollmeier, 2005). This has led to the implementation of various paediatric speech testing measures that consider both vocabulary level and test length (Neumann et al., 2012; Willberg et al., 2020).

1.7.1 Kendall Toy Test

For NZ paediatric populations, it is common practice to employ the Kendall Toy Test (KTT) to assess speech perception. The KTT has 15 images of monosyllabic nouns: ten test items and five distractor items. The child is firstly familiarized with the images, after which the clinician reads out each item, following a carrier phrase such as "point to the ___" (e.g., point to the duck), all while covering their mouth to avoid visual cues. The first 5 distractor items are presented initially at a normal volume speech level of around 55-60 dB A. For the testing items, the audiologist will typically present at below 40 dB A, as a child with NH will achieve greater than or equal to 90% correct at 35 dB A (A weighted decibels; Ministry of Health (MOH), 2016, p. 64).

Despite the wide administration of the KTT in New Zealand, several factors reduce the test's validity. First, the KTT is administered using monitored live voice that has poorer accuracy and reliability compared with recorded stimuli, as the result can be influenced by who, and where, the test is conducted (Ostergard, 1983; Theunissen et al., 2009; Uhler et al., 2016). As a result, a child's speech recognition ability is commonly overestimated (Madell & Flexer, 2014). Furthermore, a sound level meter (SLM) must be used to measure the intensity of the stimuli. The New Zealand Universal Neonatal Hearing Screening and Early Intervention Programme (UNHSEIP) states that there should be equal distance between the SLM, the child, and the clinician, (in a triangle-like arrangement) (MOH, 2016, p.64). However, this distance is frequently based of visual estimate alone which can reduce inter and intra- test reliability.

Second, there is no formal manual in New Zealand for administering or scoring, nor any normative data for the KTT; the current method was developed from the Australian test version which employed 5 vowel pairs and 5 distractor items (Antognelli, 1986). The current passing level of 35 dB A was also based on normal average pure tone thresholds ≤ 15 dB HL (Antognelli, 1986; MOH, 2016). However, hearing screening is conducted to 20 dB HL in NZ, and a pass level of 40 dB A is recommended for the KTT in NZ settings (MOH, 2016). However, there is a lack of literature appraising the validity of a pass level at 40 dB A for the KTT, raising concerns about its validity as a speech perception measure. Moreover, there are a range of techniques employed to establish the total percentage when the intensity is lifted above 35 dB A.

Third, the test uses single-word stimuli, and is conducted in quiet environments, reducing the degree to which the results represent the child's ability to communicate in the real-world. Sentences are frequently considered superior than single-word stimuli as they afford a more accurate depiction of real-life listening environments, have better sensitivity due to steeper psychometric functions, and they can assess how well a child can fill in the gaps that can indicate their communication capacity in day-to-day settings (Bell & Wilson, 2001; Madell & Flexer, 2014; Neumann et al., 2012). Using more words per trial, in the case of sentences, also improves the tests reliability which is an important factor for paediatric populations for whom the length of the test is primarily influenced by the child's fatigue level (Neumann et al., 2012).

1.7.2 Current Paediatric Speech in Noise Measures

Few measures exist to test paediatric speech recognition in noise (Schafer, 2010). Two more common tests are the Hearing in Noise Test for Children (HINT-C; Nilsson et al., 1994), and the Bamford-Kowal-Bench Speech-in-Noise test (BKB-SIN; Etymōtic Research, 2005). However, both require vocabulary levels that exceed that of a typical 5-year old child. While the BKB-SIN uses an adaptive procedure to measure the SNR loss in multi-talker babble, the use of standard SNRs make it prone to floor and ceiling affects and the results can be difficult to interpret (Etymōtic Research, 2005; Schafer, 2010). The HINT-C uses stimuli appropriate for children between six and twelve-years-old and can be conducted both in quiet and with speech-shaped noise (Nilsson et al., 1994). However, speech-shaped noise is not as representative or challenging as multi-talker babble in a classroom setting (Schafer, 2010; Sperry et al., 1997). The HINT-C is also an expensive test to run (Schafer, 2010).

A speech-in-noise test designed for younger children and currently used in NZ is the Paediatric Speech Intelligibility test (PSI; Jerger & Jerger, 1982). It is a closed-set measure, requiring that children listen to the speech stimuli and point to their response displayed on a picture card. Both monosyllabic words and sentences are used and are each presented under both quiet and noisy conditions (Jerger & Jerger, 1982). However, it employs fixed signal levels and can cause floor and ceiling effects, unless administration of multiple lists is undertaken to identify the best SNR for each child (Jerger & Jerger, 1982; Schafer, 2010). At this age, young children may not be equipped with attention spans required to perform various PSI lists at differing SNRs. Also, while the chosen masker is single talker competing noise, Schafer (2010) discusses that this may not be sufficiently challenging or representative of true noise encountered in typical classroom environments. Due to the lack of validity and reliability pertaining to the KTT, as well as various limitations pertaining to the currently available paediatric speech-in-noise measures, a new standardised paediatric sentence-based speech recognition test, with greater reliability and validity is justified.

1.7.3 Paediatric MSTs

As discussed in section 1.6, children have, on average, poorer WMC that may hinder their ability to hold onto their answer in their short-term auditory memory while they search for the right button (Cervera et al., 2009; Theunissen et al., 2009; van Rooij & Plomp, 1990). Hancock and colleagues (2007) state that more complex cognitive tasks necessitate greater processing than simple tasks which lead to longer reaction times. As an MST is a sentence test that is typically presented in noise, investigation into the impact that auditory WMC has on the test is warranted. In particular, whether a shorter sentence stimulus is more suitable for those with poorer auditory memory.

Some modified versions of the conventional Hagerman (1982) five by ten-word matrix sentence test have been developed for paediatric use including: the German Olderburger Kinder-

Satztest (OlKiSa; Neumann et al., 2012), adapted from the Oldenburge Staztest (OLSa; Wagener et al., 1999a; Wagener, et al., 1999b; Wagener, et al., 1999c); the Polish Paediatric Matrix Sentence Test (PPMST; Ozimek et al., 2012), adapted from the Polish Matrix Sentence Test (Ozimek et al., 2010); and more recently, the Finnish Simplified Matrix Sentence Test (FINSIMAT; Willberg et al., 2020) adapted from the Finnish Matrix Sentence Test (FMST; Dietz et al., 2014; 2015) and the Simplified Italian Matrix Test (SilMax; Puglisi et al., 2021) adapted from the Italian Matrix Sentence Test (ITAMatrix test; Puglisi et al., 2015). These paediatric versions use three-word pseudo-sentences consisting of a number, adjective, and object, compared to the conventional five-word sentences. Reducing the number of test items per sentence simultaneously increases the measure's sensitivity and reliability as the length of the test is dictated by the child's level of fatigue (Neumann et al., 2012). In these paediatric MSTs, children have the option to repeat what they heard in an open-set response format, or, in the case of the PPMST, a picture-pointing response version was also available for children who may have difficulty with verbal responses (Ozimek et al., 2012). Despite having fewer words per presentation, test-retest reliability is comparable (Willberg et al., 2020). Finally, the shorter sentences of the paediatric MSTs are likely to be more appropriate for the WMC of older populations.

1.8 The UCAMST-P: A New Zealand Paediatric Matrix Sentence Test

1.8.1 Development of the UCAMST-P

Jenkins-Foreman (2018) edited the original five by ten word UCAMST matrix to create a smaller three by six word matrix, the UCAMST-P (Figure 4). The first two columns of the UCAMST matrix (name and verb columns) were removed to allow the creation of three-word pseudo-sentences (quantity, adjective, object). Further removal of words to reduce the number of rows from ten to six was based on four main criteria (Jenkins-Foreman, 2018): the word's naturalness following auditory and visual editing, the word's lexical appropriateness for paediatric populations, the steepness of the psychometric slope, and the word's capability to be presented in both constant and babble noise.

Quantity	**Adjective**	**Object**
two	big	bikes
three	green	books
eight	new	hats
nine	old	shoes
ten	red	spoons
twelve	small	toys

Figure 4. The UCAMST-P matrix. *From Jenkins-Foreman (2018, p. 64).*

The UCAMST-P matrix fulfils the need to have a test that is not biased by a child's vocabulary level or higher-level language abilities that could occur with the greater cognitive demands of the original 5x10 Hagerman (1982) tests (Kosky & Boothroyd, 2003; Neumann et al., 2012). The shorter three-word pseudo-sentences have also been found to decrease the effects that attention and fatigue have on longer sentences (Neumann et al., 2012; Wagener et al., 1999a-c; Wagener & Kollmeier, 2005)

1.8.2 Response Formats

Like the UCAMST, the UCAMST-P can be administered in either a closed-set or open-set response format. A closed set may be advantageous for young paediatric populations if the child is shy or has difficulty with pronunciation, as they can compare their answer with defined alternatives, whereas open-set testing leans on their vocabulary and lexical memory (Clopper et al., 2006). However, Madell and Flexer (2014), recommend that open-set testing should be used as soon as a child is capable of the task as it affords a more realistic view of the listener's speech perception abilities with everyday communication. Ripberger (2018) also suggested that older populations with cognitive deterioration may score better in the open set, due to greater intellectual burden that is associated with self-scoring in the closed set modality, although this has not been researched further.

1.8.3 Sentence List Equivalence and Normalisation

Jenkins-Foreman (2018) assessed the UCAMST-P for sentence list equivalence within each presentation mode, equivalence between each presentation mode, and whether practice before the list made a difference to each test condition. The UCAMST-P lists were equivalent with respect to SRT and slope under the AV modality, regardless of the response format. Conversely, the SRTs were inequivalent for both response formats under the AA modality. While the slope of the intelligibility function for the AA, open set was also inequivalent, the slope for the AA, closed-set condition was equivalent. Finally, no significant impact of training was identified for the UCAMST-P with regards to the SRT or slope. Unfortunately, an error was made in calculations by Jenkins-Foreman (2018) pertaining to the UCAMST-P leading to error in this analyses of sentence list equivalence, condition equivalence, and the effect of training preceding testing (Jenkins-Foreman, 2018). This error will be remediated and reanalysed in the current thesis.

Following the generation of Jenkins-Foreman's (2018) data, Lay (2019) formulated new sentence lists to improve list equivalency in the AA, open set condition. However, these lists were based off the erroneous results produced by Jenkins-Foreman (2018) which subsequently impacted the results. The lists generated by Lay (2019) will be recalculated in the current study and compared with the amended results of Jenkins-Foreman (2018) to determine whether they are more equivalent.

To allow for interpreting UCAMST-P test scores with greater confidence, Lay (2019) also collected normative data for six to 12-year-olds in the AA, open-set condition. SRT scores improved with age up to approximately 10 years, before plateauing for the remaining age groups. This aligns with trends in current literature demonstrating a maturational effect between age and performance on auditory listening tasks speech recognition test performance (Gathercole et al., 2004; McGaffin, 2007; Neumann et al., 2012; O'Beirne et al. 2012; Wilson et al., 2010; Yau, 2018). However, there were some limitations that affected the accuracy of the normative data produced, largely pertaining to the small sample size, multiple testers, and school testing environment (Lay, 2019).

1.8.4 Cognitive Load of Current Paediatric MSTs

In the German paediatric MST (OlKiSa), children in their first year of primary school appeared to do worse than those in older year groups, requiring 1-2 dB greater SNR for equivalent performance (Neumann et al., 2012). This was attributed to an age effect whereby younger children have a poorer auditory memory span (Neumann et al. 2012). This aligns with previous research by Gathercole and colleagues (2004) who found that a child's auditory working memory, and therefore their performance on auditory listening tasks, progressively improves with age.

Additionally, Willberg and colleagues (2020) discussed that while these MSTs are intended for paediatric populations, their concise characteristics, multiple response options, and different modalities, could make them highly appropriate for elderly clients whom may have age-related cognitive decline. Failure to consider the effects of auditory memory may lead to poorer rehabilitation outcomes as the impact of working memory may augment an apparent hearing impairment, potentially deceiving the clinician and resulting in the selection of a less optimal rehabilitation strategy.

1.9 Response Times and Response Orders

1.9.1 WMC and Response Times

More complex cognitive tasks like sentence testing in noise necessitate greater processing and are more taxing on auditory working memory than simple tasks. This can result in both longer reaction times (Hancock et al., 2007; Heinrich et al., 2015; Kahana & Loftus, 1999; McGarrigle at al., 2014) and increased error rate (Kahana & Loftus, 1999). Reaction times and accuracy have been measured in previous studies of cognition to evaluate WMC (Aronen et al., 2005; Hancock et al., 2007; Hülür et al., 2019; Meister et al., 2018). Meister and colleagues (2018) examined response times of a five-word MST (the OLSA; Wagener et al., 1999a; Wagener, et al., 1999b; Wagener, et al., 1999c) as a potential measure of cognitive load during standard speech audiometry. They found that response times were faster for the higher intelligibility level (95%) compared with the lower intelligibility level (80%), and for fluctuating noise compared with stationary noise. There was also a difference between listener groups:

response times were fastest for the younger group with NH (median age 21.5 years), followed by the older group with NH (median age of 71.0 years), with the slowest responses coming from the older group with HI (median age of 74.8 years). It is expected therefore, that the shortened UCAMST-P will be less taxing on WMC, thus yielding faster and more accurate responses. In such case, it may be deemed more suitable for paediatric and older populations.

1.9.2 Effects of Stimuli on Response Time

In tests of free recall, memory span for verbal stimuli, such as digits, letters, and words, depends on two factors: the phonological complexity of the content (the number of phonemes and syllables) and on the lexical difficulty of the stimuli (Hulme et al., 1995; Poirier & Saint-Aubin, 1996; Service, 1998). As the selection of UCAMST items for the UCAMST-P was weighted towards simple vocabulary in order to be applicable for paediatric populations (Jenkins-Foreman, 2018; Trounson, 2012), this should reduce taxes on WMC. For example, some words like “dark” and “shirt” were excluded from the UCAMST-P matrix based on their lexical difficulty (Jenkins-Foreman, 2018).

Additionally, more words can be remembered in a test of free recall when the words have shorter spoken duration (the word-length effect; Baddeley et al., 1975) and when their speech sounds are similar to each other (phonological similarity effect; Conrad & Hull, 1964). Consequently, the short word, single-syllable stimuli of the three-word pseudo-sentences may be less cognitively taxing than the five-word sentences of the UCAMST, all of which have two syllables in their first word (the name item). However, the findings from such free recall tests cannot be inferred for MSTs, as tests of free recall are do not use sentence-stimuli, nor are they administered with background noise. Nonetheless, it was expected that response times would be faster for the easier condition (3-word UCAMST-P, higher SNR, practice) compared with the harder conditions (5-word UCAMST, lower SNR, no practice) due to the assumed differences in their cognitive demands. It was anticipated that following the current research, it may therefore be appropriate to recommend the UCAMST-P for both paediatric and elderly populations to account for their reduced WMC.

1.9.3 Response Order

Alongside response times, the order with which participants select their response may provide complementary insight into how the UCAMST and the UCAMST-P are influenced by a listener's WMC. When testing memory under free-recall, items presented earlier and later are recalled more accurately than items presented midway through. The advantage given to the first and last words are deemed primacy and recency effects, respectively (Kahana & Loftus, 1999). Participants who are required to recall words, rehearse the words in their working memory as they are presented. Specifically, they rehearse the previous words with each new word that is added (Glenberg et al., 1980; Marshall & Werder, 1972; Rundus, 1971, 1980). Due to the length of the lists that are recalled in those tests, Murdock (1962) suggested this process may incorporate both long-term memory for the first few items, and short-term memory for the final few. However, these traditional tests of free recall employ words from a large set that greatly exceeds the recall required for that of the 3-word and 5-word MSTs. It is unclear whether these findings of a free-recall test align with the semantically unpredictable sentences of the UCAMST and UCAMST-P, particularly when tested under degraded listening conditions where working memory is heavily taxed to make sense of the signal (Akeroyd, 2008; Souza et al., 2015).

Another key notion in memory research is that stimuli which are processed sequentially, end up associated with each other. For example, after recalling item 1, item 2 is more likely to be immediately recalled than item 5. The chance of forward sequential transitions (item 2 then item

3) is approximately double that of backward transitions (item 5 then item 4) (Kahana & Loftus, 1999). As Kahana and Loftus (1999) suggest that sequential responses are twice as likely in tasks of free recall under simple listening conditions, it is expected that the responses for the UCAMST-P and tests under high SNR conditions will be quick and sequential as the responses can be held in the short-term memory, requiring just bottom-up processing (Pichora-Fuller et al., 1995). On the other hand, as difficult listening situations like the 5-word UCAMST and a reduced SNR are likely more taxing on WMC (Akeroyd, 2008; Souza et al., 2015) responses may be both slower and less sequential (Beechey et al., 2020; Rönnberg et al., 2013). In this case the UCAMST-P could be recommended as the more appropriate speech recognition test for individuals with reduced WMC.

1.10 Study Rationale

1.10.1 Part A: Revised Sentence Equivalence Results

Comprehensive research has been undertaken to develop and prepare the UCAMST and UCASMT-P for clinical use (Jenkins-Foreman, 2018; Lay, 2019; McClelland, 2015; O'Beirne, et al., 2015; Ripberger, 2018; Stone, 2016; Trounson, 2012). However, due to the error made in the data calculations associated with the UCAMST-P in Jenkins-Foreman's (2018) thesis, re-calculation and re-analysis of the results associated with the UCAMST-P is required to satisfy the original research aims; to assess the reliability and sensitivity of the newly produced UCAMST-P in estimating SRTs.

1.10.2 Part B: Auditory Memory and the UCAMST

The clear effect of auditory WMC on speech recognition testing necessitates research into whether the UCAMST-P is more suitable than the UCAMST for populations with poorer auditory working memory. Failure to consider the effects of WMC may lead to poorer rehabilitation outcomes as the impact of working memory may augment an apparent hearing impairment, potentially deceiving the clinician into selecting a less than optimal rehabilitation strategy. As discussed above in Section 1.6, both children and the elderly have shorter WMC (Cervera et al., 2009; Theunissen et al., 2009; van Rooij & Plomp, 1990) which may result in difficulty holding their answer in their working memory while they search for the right button.

Unanalysed data from Jenkins-Foreman (2018) includes participants reaction times taken to select their response on the matrix response panel for both the UCAMST and UCAMST-P under the closed-set condition, as well as whether the response was correct, whether the stimuli was presented with a high or low SNR (roughly corresponding to the 80% and 20% intelligibility levels), and whether or not the sentence was preceded by practice. It was expected that the harder conditions (5-word, high SNR, and without practice), would yield longer response times than the easier conditions (3-word, low SNR, and with practice) due to the extra demands on auditory memory. The order with which words are pressed in each sentence were also collected by Jenkins-Foreman (2018). This information affords insight into how auditory memory affects the responses of listeners. Similarly, it was expected that the harder conditions would give rise to less sequential responses compared with the easier conditions, as the more difficult listening situations requiring greater listening effort might cause listeners to “lock in” those responses they are confident in before attempting those they might have less confidence in.

1.11 Aims and Hypotheses

1.11.1 Part A: Revised List Equivalence

The aim of the current study was to recalculate and discuss the data sourced from Jenkins-Foreman (2018) pertaining to the AA condition of the UCAMST-P. To do so, similar research questions to those proposed by Jenkins-Foreman (2018; p. 47-48) will be revisited:

1) Are the lists equivalent within each condition (i.e. AA, open-set; AA, closed-set) with regards to SRT and slope for the UCAMST-P?
2) Are the open- and closed-set response formats equivalent within each mode of presentation (i.e. AA, open-set; AA, closed-set) with regards to SRT and slope for the UCAMST-P?

The following hypotheses were initially proposed by Jenkins-Foreman (2018; p. 48-51) and will be revisited for the current research project:

For research question 1:

1) That no significant differences would be found between the UCAMST-P sentence lists with regards to <u>SRT</u> in the:
    a. AA, open-set condition
    b. AA, closed-set condition
2) That no significant differences would be found between the UCAMST-P sentence lists with regards to <u>slope</u> in the:
    a. AA, open-set condition
    b. AA, closed-set condition

For research question 2:

3) That no significant differences would be found between the open-set and closed-set response formats of the UCAMST-P in the AA mode of presentation with regards to:
    a. SRT
    b. Slope

1.11.2 Part B: Response Times and Auditory Memory

This thesis used data from Jenkins-Foreman (2018) to examine whether the UCAMST-P may be more suitable for those with reduced or developing working memory, including NZ's paediatric population and senior adults. In order to evaluate the impact of auditory memory, the current research project endeavoured to answer the following research questions:

3) Is there a significant difference between the tests and conditions, (i.e. UCAMST-P, UCAMST; high SNR, low SNR; practice, no practice) with regards to response times?
4) Is there a significant difference between the tests and conditions, (i.e. UCAMST-P, UCAMST; high SNR, low SNR; practice, no practice) with regards to word selection order?

The following hypotheses were proposed for the current research project:

For research question 3:

4) That response times will be faster for the:
    a. UCAMST-P than the UCAMST
    b. Higher SNR than poorer SNR
    c. With practice than without practice

For research question 4:

5) That button pressing will be more sequential for the:

    a. UCAMST-P than the UCAMST

    b. Higher SNR than poorer SNR

    c. With practice than without practice

Chapter Two: *Methods*

2.1 Data Collection

The current study will reanalyse Jenkins-Foreman's (2018) results pertaining to the UCAMST-P due to a miscalculation in the original analysis, as well as analyse surplus data from Jenkins-Foreman (2018) regarding the reaction times of participants for the UCAMST and UCAMST-P using the AA modality under closed-set conditions. Importantly, use of this pre-existing data means that data collection was performed wholly by Jenkins-Foreman (2018).

2.1.1 Participants (Jenkins-Foreman, 2018)

2.1.1.1 Inclusion Criteria

Participants had to be native NZ English speakers to verify the appropriate use of the UCAMST-P in NZ settings, particularly as speech intelligibility is significantly reduced when listening to a non-native speaker (Zokoll et al., 2013). Participants also needed to be at least 18 years old as a person's ability to continually focus for a given task improves during adolescence (Betts et al., 2006). Finally, they were required to have NH (in accordance with Goodman (1965)) so that data was not confounded by a HI, as well as sufficient dexterity and visual acuity to select words on a screen.

2.1.1.2 Participants

Participants were recruited through the University of Canterbury Department of Communication Disorders. An email invitation was circulated that outlined the research project and the inclusion criteria for participants. To achieve sufficient statistical power based on G*Power 3.1 calculations (Jenkins-Foreman, 2018), 31 participants were needed, although, 40 participants were required based on the testing structure. A total of 43 participants aged between

19 to 48 years old met the inclusion criteria and were recruited. 11 were males and 32 were females. A $20 voucher was given to all participants to thank them for their contribution.

2.1.2 Stimuli

2.1.2.1 UCAMST

Testing all 100,000 potential UCAMST sentences was unfeasible so an iterative procedure (described by Stone, 2016) was instead used to create 16 lists of 10 sentences (appendix A1). Every list had precisely one occurrence of each word in every position so that each list was constrained to the same average slope value. Within each list, care was taken to homogenise the slope values so that listeners did not encounter some sentences with a very low slope and others with a very high slope. Each of the lists had a very low standard deviation of slope values, and no sentences were repeated between lists.

2.1.2.2 UCAMST-P

The sentences for the UCAMST-P were produced by Jenkins-Foreman (2018) who systematically created all 216 (6^3) available pseudo-sentences and then rejected sentences with a transition judder[1] magnitude of three or greater (Stone, 2016; Trounson, 2012). This process cut the available number of pseudo-sentences from 216 to 162. All 162 pseudo-sentences could be analysed in the same time frame given for the 16 lists of 10 sentences from the UCAMST's 5 by 10 matrix, in an almost identical procedure. Consideration was made to avoid any replicate two-word pairs in a single list (e.g. no repeats of "ten small" or "three books") and to avoid any noticeable patterns in responses (e.g. third word in each column). This resulted in 162 pseudo-

---

[1] A "judder" is an interruption to the smooth visual transition between two words in a matrix sentence when presented in the auditory-visual condition. Sentences and pseudo-sentences with a high degree of measured judder were excluded from the test lists.

sentences being sorted into 14 lists of 10 and two lists of 11 for testing (appendix A2). The two leftover sentences “twelve old bikes” and “ten new spoons” were manually distributed into lists.

2.1.3 Presentation Levels

Sentence stimuli were presented bilaterally at 65 dB SPL. To every SNR that was acquired, 3.85 dB SPL was added in accordance with the signal calibrations found by Stone (2016). Although the UCAMST and UCAMST-P use an adaptive tracking procedure to locate the SRT, this study used set SNR levels for the verification procedure. All sentence lists in the UCAMST and the UCAMST-P were presented in constant masking noise at two different SNRs for both the open-set (-11.6 dB SNR and -6.0 dB SNR) and closed-set (-14.0 dB SNR and -7.4 dB SNR) response formats to allow for approximation of the pair of compromise (the location at which 80% and 20% scores are predicted to lie). This allowed simultaneous estimates of the slope and the psychometric function to be found, from which the SRT could also be derived (Brand & Kollmeier, 2002). Each SNR was randomly distributed to half of the sentences in each list to ensure that an equal proportion of sentences were presented at both SNRs.

2.1.4 Experimental Instrumentation

The initial hearing screening and experimental procedures were conducted at the University of Canterbury Speech and Hearing Clinic (Christchurch, New Zealand) in sound-proof audiological testing booths. The preliminary hearing screening was conducted in accordance with the NZAS best practice guidelines to obtain participant’s audiometric thresholds (NZAS, 2016). This required testing of octave pure tones between 250 to 8000 Hz using a calibrated Grason-Stadler GSE 61 clinical audiometer. Tones were administered via Telephonics TDH-50P supra-aural headphones, and a RadioEar B-71 bone transducer was used to obtain bone conduction thresholds using pure-tones at 500, 1000, 2000, and 4000 Hz. Participants responded by pressing the response button connected to the audiometer. The UCAMST and UCAMST-P software that was used for normalisation was developed by Professor Greg O'Beirne using LabVIEW. An HP EliteDesk 800 G1 and Phillips Brilliance 241B monitor were used to run the software. These were connected to an ēlo touch-sensitive monitor (ēlo ET17115L, Tyco Electronics, CA, USA) which displayed the visual modality and presented response options for the participants in the closed-set response formats for both the UCAMST and UCAMST-P. The sentence stimuli and the masking noise were both presented via Senheiser HD 280 Pro (64 Ω impedance) circumaural headphones driven by a Sound Blaster X-Fi Surround 5.1 Pro USB sound card connected to the computer.

2.1.5 Experimental Procedures

Each participant read information and consent forms and were given time to ask questions, after which each participant gave informed consent. Participants were asked questions about their ear health and otoscopy was to check the health of the outer and middle ear. A hearing screening was performed to ensure that participants had normal hearing so as not to confound the test outcomes. All participants had normal hearing and could proceed with the study (appendix B).

Each participant was then randomly assigned to one of four testing blocks, each block presenting the tests and presentation modes in different orders (Table 1; presented from left to right). In total, the participants were presented with 40 sentences for each of the five-word and three-word sentences, under both the AA and AV presentation modes. The participants responded in both open-set and then closed-set formats where each constituent word was scored as correct or incorrect. Jenkins-Foreman (2018) administered the open-set response format before the closed-set format to reduce the likelihood of any learning effects that may occur if the closed-set responses were conducted first. In this way, the block testing method incorporated every test condition so that each participant could act as their own control**.**

Table 1. Block Testing Conditions

Block One											
Auditory-alone						**Auditory-visual**					
Open-set			Closed-set			Open-set			Closed-set		
P5	5	3	P5	5	3	P5	5	3	P5	5	3
Block Two											
Auditory-visual						**Auditory-alone**					
Open-set			Closed-set			Open-set			Closed-set		
P5	5	3	P5	5	3	P5	5	3	P5	5	3
Block Three											
Auditory-alone						**Auditory-visual**					
Open-set			Closed-set			Open-set			Closed-set		
P3	3	5	P3	3	5	P3	3	5	P3	3	5
Block Four											
Auditory-visual						**Auditory-alone**					
Open-set			Closed-set			Open-set			Closed-set		
P3	3	5	P3	3	5	P3	3	5	P3	3	5

Note. Auditory-alone = auditory-alone mode of presentation; Auditory-visual = Auditory-visual mode of presentation; Open-set = open-set response format; Closed-set = closed-set response format; P = practice; 5 = five-word sentences from the UCAMST; 3 = three-word sentences from the UCAMST-P.

Table from Jenkins-Foreman (2018, p. 62).

All experimental testing was completed in a sound-treated booth. For the test, participants were required to wear headphones which administered the UCAMST tests binaurally. Participants had a touch screen display in front of them that displayed the video for the auditory-visual presentations when required, and then showed the closed-set testing matrices when testing in that mode; the 50-word response table for the UCAMST, or the 18-word response table for the UCAMST-P (Figures 5 and 6).

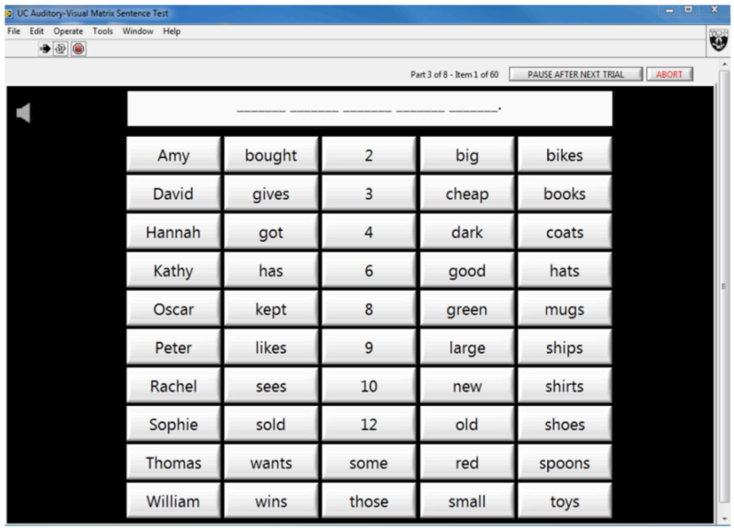


Figure 5. The UCAMST closed-set response panel. *Image retrieved from Jenkins-Foreman (2018, p.63).*

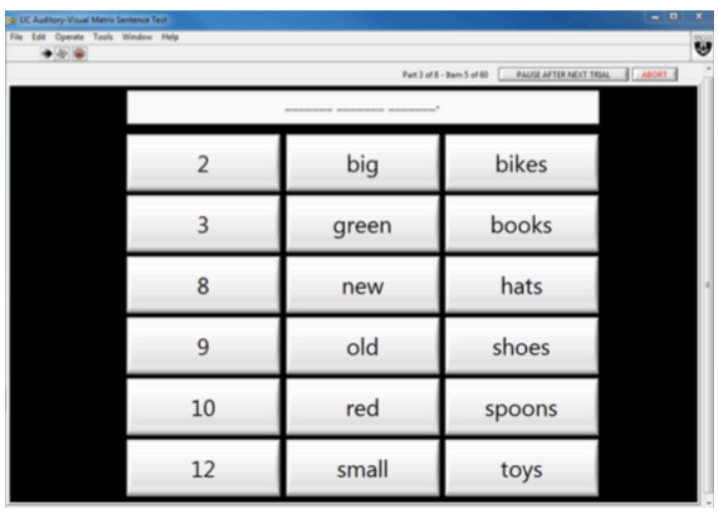


Figure 6. The UCAMST-P closed-set response panel. *Image retrieved from Jenkins-Foreman (2018, p.64).*

Participants were instructed verbally that they would hear multiple sentences with differing intensities of background noise. Although written instructions describing each task were provided on the touch-sensitive monitor at the beginning of each new test, the participants were also verbally instructed to repeat the sentence out loud in the case of open-set testing, or to select their answer on the touch screen when this option was available. Guessing was encouraged if they were not sure. Participants were also notified that a full sentence response was required to move on to the next sentence in the closed-set modality.

At the beginning of each presentation mode and response format, Jenkins-Foreman (2018) presented two practice lists of 20 sentences to make sure that the participants understood the task and to allow for the training effect, as well as to stabilise their performance before starting to test (Wagener et al., 2003). Including the practice lists, this summed up to 480 sentences. Notably, the participants were encouraged to take regular breaks as necessary as the test was long and required sustained concentration. The testing procedure took approximately 80 minutes per participant, excluding times for breaks

2.1.6 Scoring Procedures

Word-based scoring methods were employed by Jenkins-Foreman, owing to the findings of McClelland (2015) who found that the UCAMST generated steeper slope scores when using word scoring compared with fragment scoring. The five-word sentences of the UCAMST were therefore, allocated five points per sentence, and the three-word pseudo-sentences of the UCAMST-P were scored out of three (Jenkins-Foreman, 2018). An average score for a UCAMST sentence at a particular SNR was derived by dividing the sum of its correct words by five. Unfortunately, an error in the custom-written analysis software written by the senior supervisor caused the average score for each three-word UCAMST-P pseudo-sentence to be calculated in the same way, dividing by five instead of by three. This error greatly reduced the apparent performance of the UCAMST-P in the 2018 Jenkins-Foreman study. Reaction times were measured in seconds from the response of their first word, to the time that each following word was pressed on the screen. The word order in which the pseudo-sentence was presented was also recorded.

2.2 Current Study

2.2.1 Experimental Instrumentation

The custom-written software that was used to condense data from Jenkins-Foreman (2018) was amended so that the error causing the average UCAMST-P results to be divided by five, rather than three, was removed. This corrected data was then subjected to statistical analysis as described in Section 2.2.3 below. The format of the data file saved for each sentence presentation is shown in Tables 2 and 3 below:

Table 2: A 5x10 matrix response from Participant 20 in the auditory-alone closed-set no-practice condition (23 Sep 2017 1:41:34 pm)

Time (s)	SNR	Presented sentence					Selected sentence					Correct?					%Words Correct	Response time (s)				
456.3	-7.43	Rachel	got	two	good	toys	Rachel	bought	two	good	toys	1	0	1	1	1	80	1.75	2.61	4.05	4.73	5.61

Table 3: A 3x6 matrix response from Participant 20 in the auditory-alone closed-set with-practice condition (23 Sep 2017 1:33:03 pm)

Time (s)	SNR	Presented sentence					Selected sentence					Correct?					%Words Correct	Response time (s)				
208.4	-7.43			nine	green	spoons			nine	red	spoons			1	0	1	67			0.65	1.91	2.47

In both of these tables, the response time for the pressing of each of the closed-set response buttons is given as seconds following the end of the played stimulus. Just as the analysis software aggregated the data regarding the proportion of correct responses for each word in each sentence, it was able to be modified to aggregate the timing data under the "Response time (s)" heading for the different test conditions.

Microsoft Excel version 16.4.3 was used to examine the aggregated data, and to produce figures pertaining to Part B. Version 26.0.0.1 of the IBM Statistical Package for the Social Sciences (SPSS) was also used to perform all statistical analyses on the data pertaining to Part A.

2.2.2 Data

For the UCAMST-P, the presence or absence of a practice round was found by Jenkins-Foreman (2018) to make no significant differences to SRT or slope for any test condition (i.e. AA, open-set; AA, closed-set; AV, open-set; AV, closed-set). For Part A of the current study, therefore, the practice and no-practice conditions of the UCAMST-P were combined to increase the number of trials and the statistical power of the analysis.

Based on findings from Jenkins-Foreman (2018), the SRT and slope values from the AV condition were incomparable with those under the AA condition as the AV slope values were much lower, regardless of the test or response format employed. Although they were included in Jenkins-Foreman's (2018) research, they have been excluded from this study. Data was further excluded on the basis that there were insufficient data points to create a realistic psychometric function.

2.2.3 Planned Statistical Analyses

For Part A, it was planned to use two separate univariate analyses of variance (ANOVA) to analyse list equivalency and condition equivalency with respect to SRT under the AA modality. Conversely, it was planned to use non-parametric tests to examine list and condition equivalency with respect to slope. For Part B, figures produced in Excel were used to allow conclusions to be drawn based on response times and response orders.

Chapter Three: *Results*

PART A

3.1 Revised UCAMST-P results

As described above, an error was observed in the previous calculation of the UCAMST-P results and subsequent analysis was therefore invalid (Jenkins-Foreman, 2018; Tables 6, 7, 8). The incorrect results stemmed from an error in calculating the average comprehension results, dividing the total by five, rather than three. Consequently, the set of equivalent sentence lists created by Lay (2019), which were made with reference to these results, may also be erroneous. The aim of the current study (Part A) was to re-conduct statistical analyses of these lists with corrected data.

As no significant differences were found in SRT or slope for when each UCAMST-P test condition (i.e. AA*,* open-set; AA, closed-set; AV, open-set; AV, closed-set) was, or was not preceded by practice (Jenkins-Foreman, 2018), the practice and no practice conditions were combined to improve the statistical power of the current study. Additionally, the results pertaining to list and condition equivalence under the AV condition were poor and thus excluded from the current study. Data was further excluded on the basis that there were insufficient data points to plot a realistic psychometric function.

The SRT and psychometric slope data for each sentence list produced by Jenkins-Foreman (2018) for the AA, open-set and AA, closed-set condition are displayed in Appendices C1 and C2. A univariate ANOVA was used to assess list equivalency and condition equivalency for SRT. However, significant bias that could violate the assumption of normality in the slope data meant that non-parametric tests had to be employed for these analyses. The results for list equivalency (research question 1), revealed no statistically significant differences between list with regards to SRT in either of the AA response formats, demonstrating that the lists are equivalent under these conditions. However, there was a significant difference in list equivalency with respect to slope under both AA response formats. Results pertaining to the effect of response formats (research question 2), showed that the open- and closed-set formats are not equivalent with respect to SRT, nor slope.

3.2 Revised UCMAST-P List Equivalence

3.2.1 Original List Equivalence

The original calculations regarding the five-word version of the UCAMST were accurate and subsequent speech intelligibility functions (shown in Figure 7) were correct. However, the erroneous calculations and resulting data in the Jenkins-Foreman (2018) thesis gave rise to inaccurate speech intelligibility functions for the UCAMST-P, which will be shown in Figure 10 on Page 54 for purposes of comparison with the corrected data.

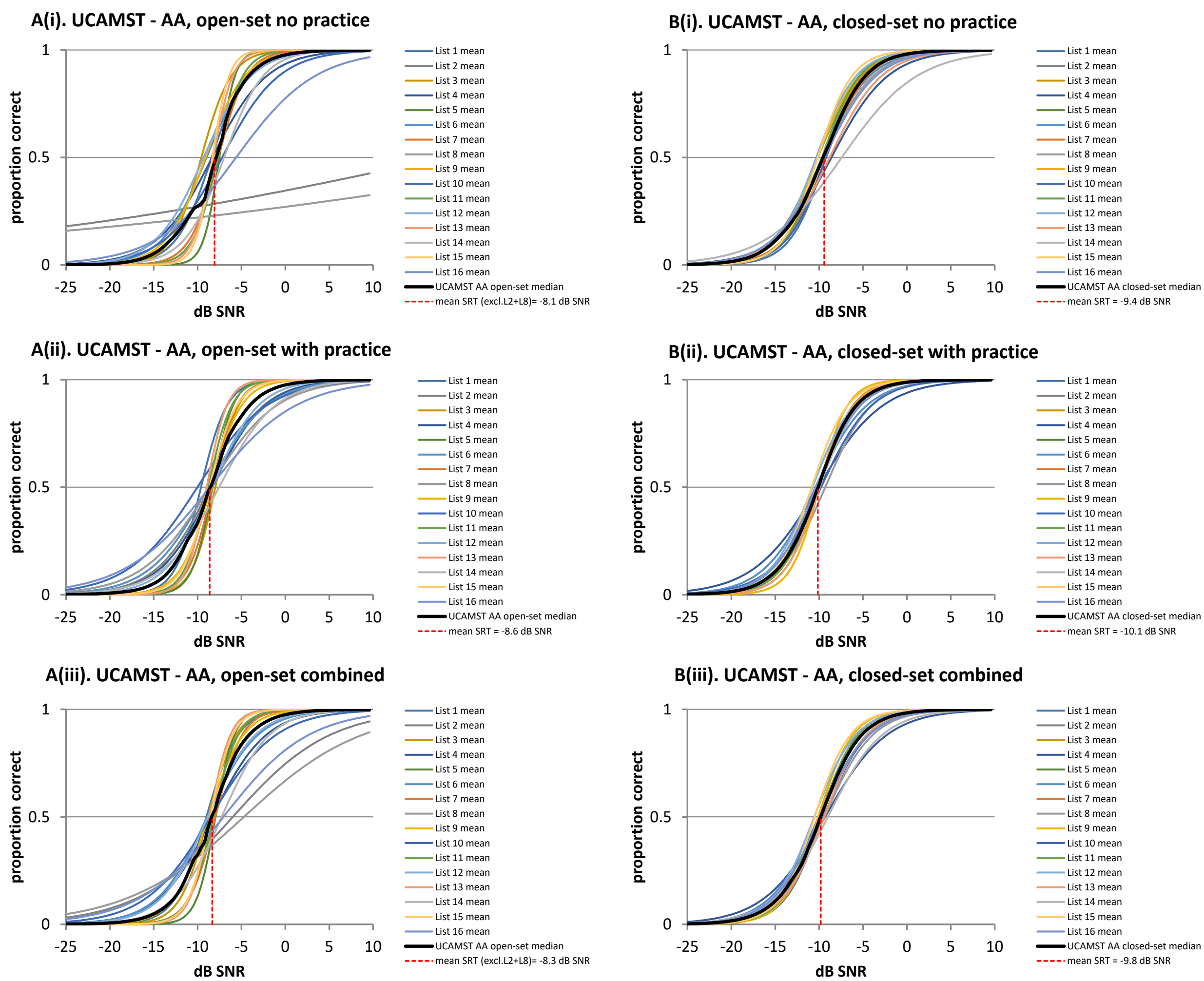


Figure 7. These panels show the original correct speech intelligibility functions for each sentence list in the AA mode of presentation for the five-word UCAMST in both the open- (A) and closed-set (B) response formats generated for i) when practice did not immediately precede the condition, ii) when the condition was preceded by practice, and iii) both of these combined.

3.2.2 Revised List Equivalence

To establish whether the lists in each condition were equivalent with regards to SRT, a univariate ANOVA was conducted (Table 4; Figures 8 and 9). For data pertaining to the slope, significant bias in the data necessitated the use of a non-parametric, Kruskal-Wallis test for list equivalence (Table 7).

Table 4. Results of the univariate ANOVA for list equivalence (SRT).

Condition	Sum of squares	df	Mean square	F	p	η_p^2
AA, Open	109	15	7.292	1.491	.116	.135
AA, Closed	26	15	1.705	.567	.896	.055

Note. df = degrees of freedom; F = F-ratio; p = p-value; η_p^2 *= partial eta-squared.*

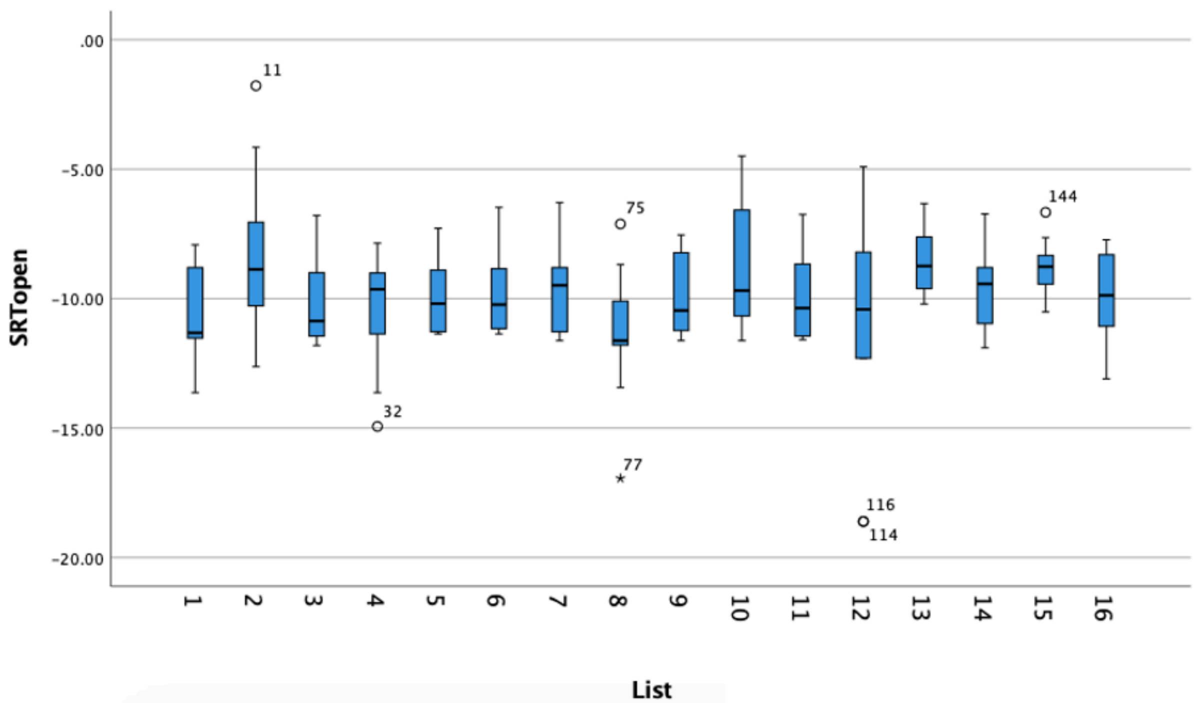


Figure 8. The SRT for each sentence list in the UCAMST-P under the AA, open-set condition.

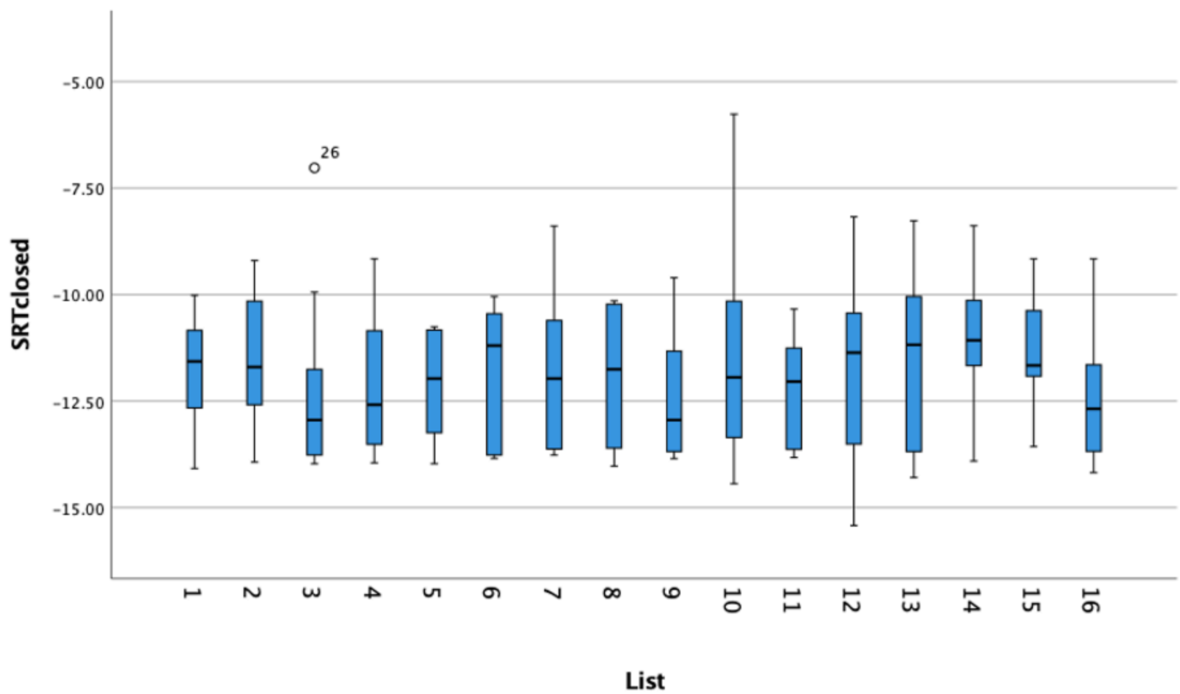


Figure 9. The SRT for each sentence list in the UCAMST-P under the AA, closed-set condition.

Hypothesis (1) - That no significant differences would be found between the UCAMST-P sentence lists with regards to SRT in the (a) AA, open-set condition; (b) AA, closed-set condition.

A univariate ANOVA performed on the UCAMST-P data revealed no statistically significant differences in SRT between sentence lists when presented in the AA, open-set condition nor between sentence lists in the AA, closed-set condition (Table 4). Hypotheses (1a) and (1b) are supported. This aligns with Lay's (2019) sentence lists: there are no significant differences between sentence lists in the AA, open-set ($F(11, 108) = 1.310$, $p = .229$, $\eta_p^2 = 0.118$) nor between sentence lists in the AA, closed-set condition ($F(11, 108) = .686$, $p = .749$, $\eta_p^2 = 0.065$) (Appendix D3).

Although there is no significant main effect, post-hoc testing was performed on SRT data for the AA, open-set and AA, closed-set conditions using the test of least significant difference

(LSD) to discern whether any sentence lists were significantly different to others (Tables 5, 6). In the AA, open-set condition, twelve significant differences are observed between lists, all of which involve lists 2, 8, 12, and 13. No significant differences are recorded between any lists under the closed-set response format.

Table 5. Pairwise comparison for SRT between AA, open-set lists

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1		**0.023**	0.721	0.931	0.541	0.492	0.432	0.450	0.506	0.079	0.503	0.601	**0.044**	0.301	0.079	0.539
2			0.054	**0.025**	0.093	0.140	0.119	**0.003**	0.104	0.593	0.105	**0.005**	0.785	0.208	0.596	0.094
3				0.780	0.799	0.725	0.673	0.271	0.758	0.161	0.754	0.380	0.097	0.497	0.159	0.797
4					0.590	0.535	0.473	0.392	0.553	0.087	0.549	0.534	**0.048**	0.331	0.086	0.588
5						0.911	0.871	0.178	0.957	0.250	0.953	0.258	0.159	0.671	0.248	0.997
6							0.969	0.167	0.951	0.330	0.955	0.239	0.223	0.772	0.328	0.914
7								0.125	0.915	0.310	0.919	0.187	0.200	0.785	0.307	0.874
8									0.162	**0.014**	0.160	0.805	**0.007**	0.079	**0.014**	0.177
9										0.273	0.996	0.236	0.175	0.710	0.271	0.960
10											0.275	**0.023**	0.793	0.468	0.996	0.252
11												0.234	0.177	0.714	0.273	0.956
12													**0.012**	0.120	**0.023**	0.256
13														0.323	0.797	0.160
14															0.465	0.673
15																0.250
16																

Note: significance ($p \leq 0.05$) indicated in bold.

Table 6. Pairwise comparison for SRT between AA, closed-set lists

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1		0.717	0.584	0.728	0.670	0.866	0.946	0.788	0.417	0.558	0.573	0.837	0.675	0.307	0.573	0.486
2			0.363	0.472	0.430	0.846	0.761	0.528	0.241	0.823	0.355	0.875	0.955	0.509	0.840	0.290
3				0.832	0.903	0.474	0.530	0.780	0.791	0.258	0.988	0.451	0.334	0.118	0.267	0.882
4					0.930	0.603	0.670	0.942	0.628	0.344	0.819	0.576	0.438	0.164	0.355	0.715
5						0.552	0.614	0.875	0.699	0.312	0.891	0.527	0.398	0.148	0.323	0.787
6							0.916	0.662	0.327	0.676	0.465	0.970	0.803	0.393	0.693	0.387
7								0.732	0.369	0.594	0.520	0.886	0.718	0.328	0.610	0.435
8									0.586	0.393	0.769	0.635	0.492	0.197	0.405	0.669
9										0.163	0.803	0.309	0.219	0.068	0.170	0.907
10											0.251	0.704	0.867	0.662	0.982	0.201
11												0.442	0.327	0.114	0.261	0.894
12													0.831	0.414	0.720	0.367
13														0.545	0.884	0.266
14															0.646	0.087
15																0.209
16																

Hypothesis (2) - That no significant differences would be found between the UCAMST-P sentence lists with regards to slope in the (a) AA, open-set condition; (b) AA, closed-set condition

A Kruskal-Wallis H test revealed significant differences in the slopes of the UCAMST-P sentence lists in both the AA, open-set and AA closed-set conditions (Table 7), thus rejecting hypotheses (2a and 2b). Similarly, the lists generated by Lay (2019; Appendix D4) are significantly different with respect to slope in the AA, open-set condition (χ^2 (11) = 39.32, p <.001). However, there is no statistically significant difference between the sentence lists generated by Lay (2019) in the AA-open set condition (χ^2 (11) = 9.45, p = 0.580).

Table 7. Results of the Kruskal-Wallis H test for list equivalence (slope).

Condition	df	χ^2	*p*
AA, Open	15	77.444	<.001
AA, Closed	15	26.837	.030

Note. df = degrees of freedom

Post-hoc analyses were not performed on the data pertaining to slope as the assumption of normality was violated due to the presence of significant bias. However, graphs were created to aid the visualisation of the similarities and differences between the average slopes of each data set. The speech intelligibility functions in the original erroneous data set are shown in Figure 10 while the corrected average slopes of the UCAMST-P sentence lists are displayed in Figure 11. A visual comparison is made in Figure 12 between the UCAMST with the UCAMST-P, as well Figure 13 which compares the sentence lists of Jenkins-Foreman (2018) with Lay (2019).

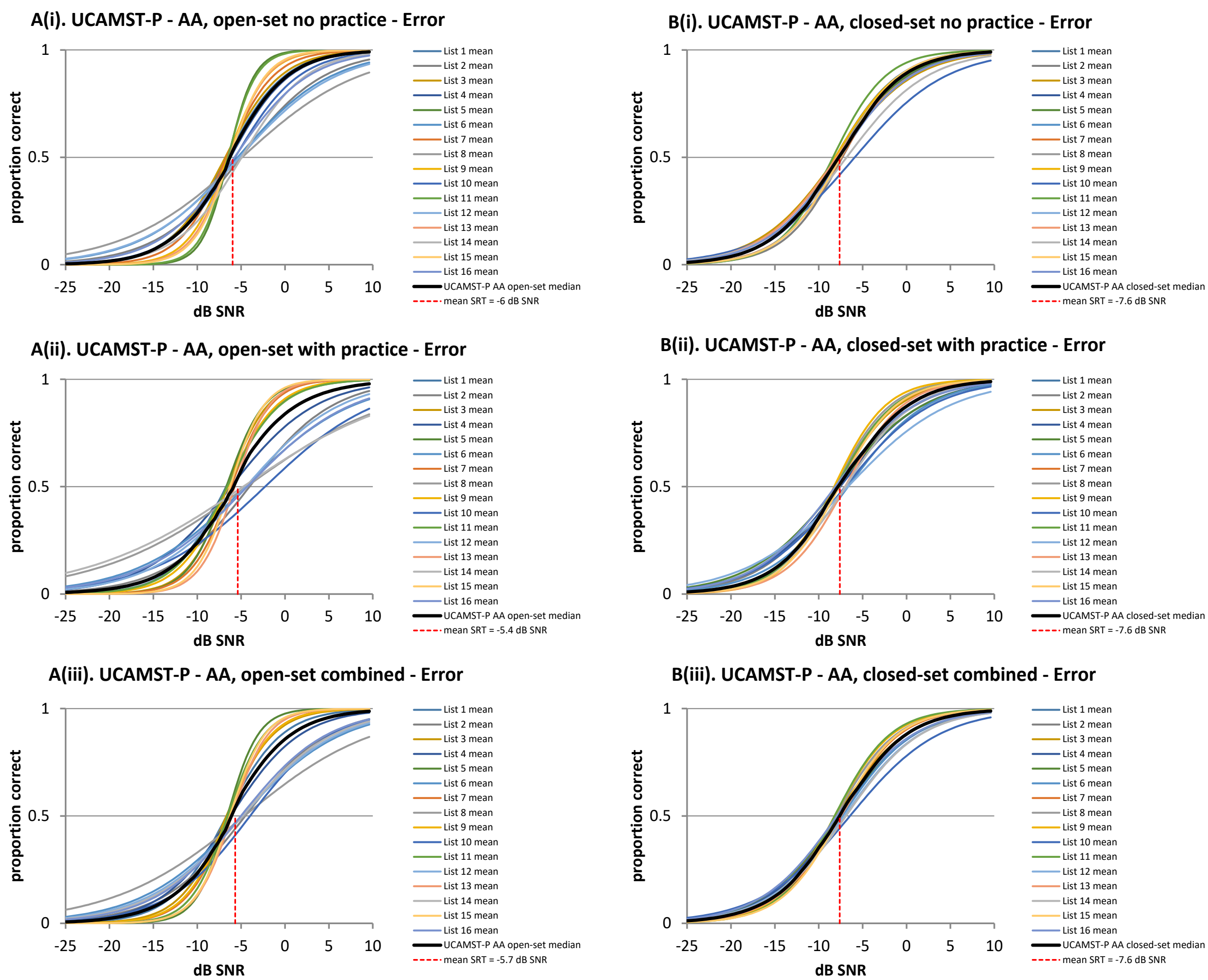


Figure 10. The original erroneous speech intelligibility functions for each sentence list in the AA mode of presentation for the three-word UCAMST-P in both the open- (A) and closed-set (B) response formats generated for i) when practice did not immediately precede the condition, ii) when the condition was preceded by practice, and iii) both of these data sets combined.

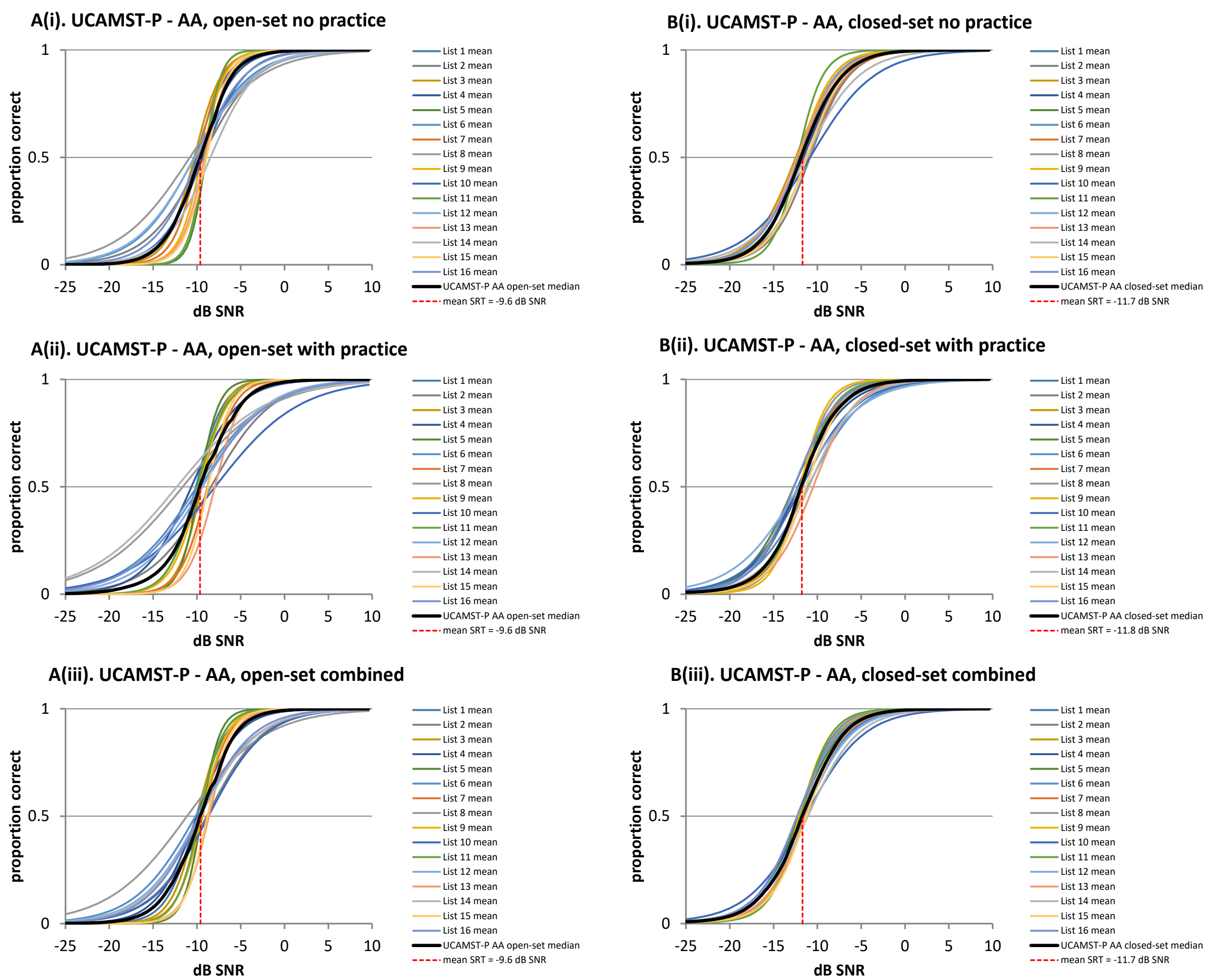


Figure 11. The corrected speech intelligibility functions for each sentence list in the AA mode of presentation for the three-word UCAMST-P in both the open- (A) and closed-set (B) response formats generated for i) when practice did not immediately precede the condition, ii) when the condition was preceded by practice, and iii) both of these combined.

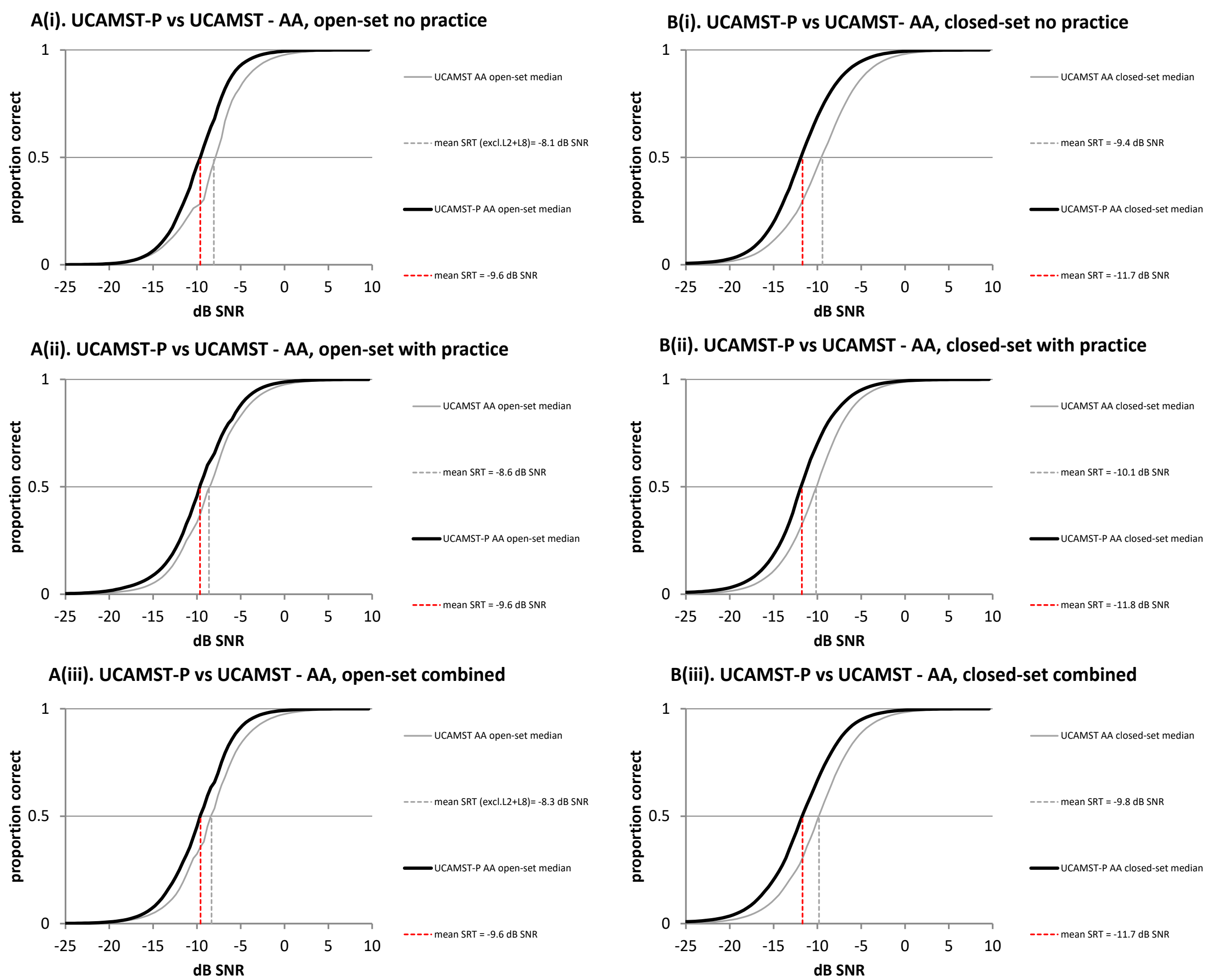


Figure 12. Intelligibility functions of the corrected UCAMST-P compared with the UCAMST in both the closed-set and open-set conditions, in the with practice, without practice, and combined with-and-without-practice conditions.

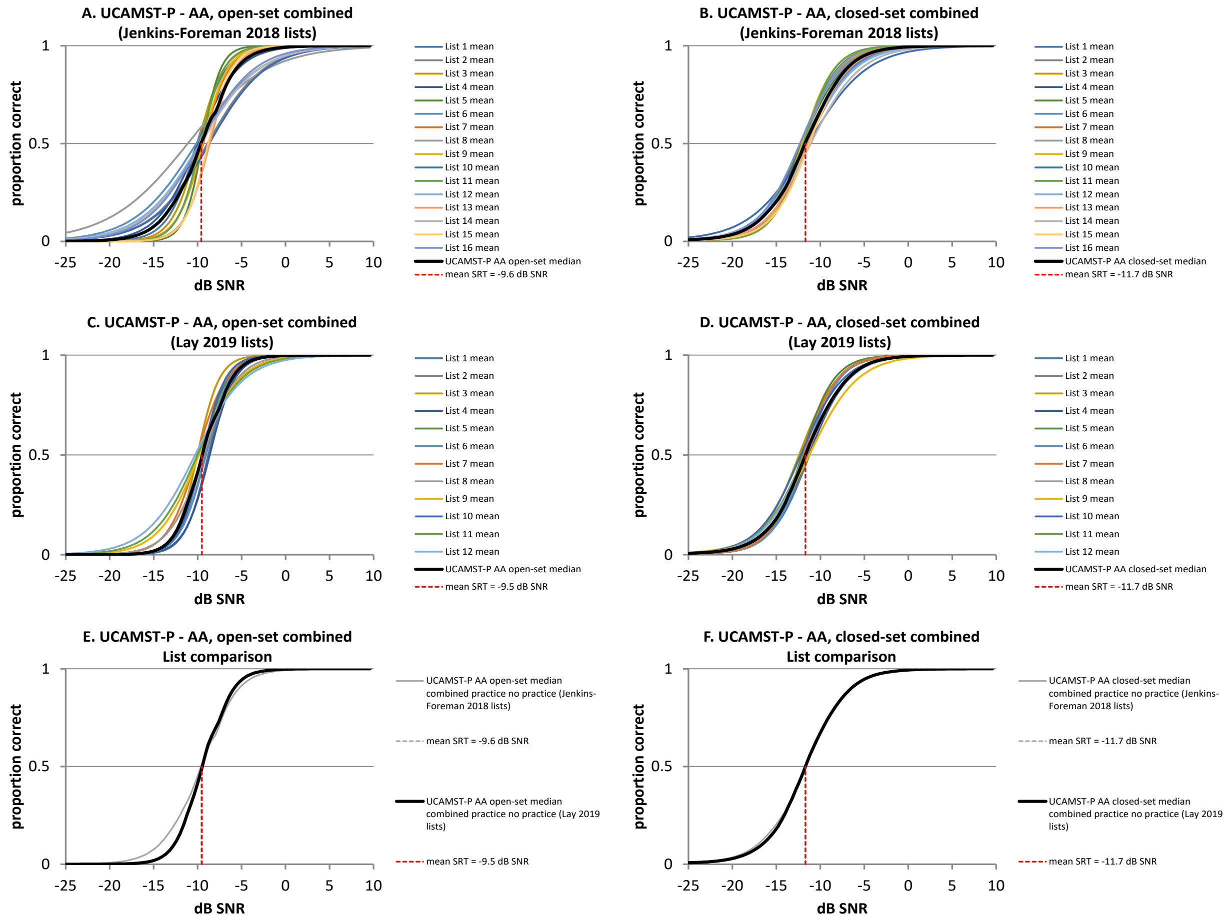


Figure 13. Intelligibility functions of Jenkins-Foreman's (2018) corrected UCAMST-P sentence lists compared with Lay's (2019) adjusted UCAMST-P sentence lists in both the open-set and closed-set conditions, combining data from the with and without practice conditions.

3.3 Revised UCAMST-P Condition Equivalence

To discern whether the open-set and closed-set conditions of the UCAMST-P are equivalent, a univariate ANOVA was performed on SRT data, while a non-parametric Mann-Whitney test was performed on the slope data. The results pertaining to hypothesis (3) are reported in tables 8 and 9, and Figure 14.

Table 8. Results of the univariate ANOVA for condition equivalence (SRT).

Condition	Sum of squares	df	Mean square	*F*	*p*	η_p^2
AA	343.643	1	343.643	87.2	<.001	.232

Note. ANOVA = Analysis of Variance; AA = auditory-alone; df = degrees of freedom; F = F-ratio; p = p-value; η_p^2 *= partial eta-squared.*

Table 9. Results of the Mann-Whitney test for condition equivalence (slope).

Condition	*U*	*p*
AA	11,222	.046

Note. U = Mann-Whitney; *p = Asymp Sig.*

Hypothesis (3) - That no significant differences would be found between the open-set and closed-set response formats of the UCAMST-P in the AA mode of presentation with regards to (a) SRT and (b) slope.

A univariate ANOVA revealed a significant difference in SRT between the AA, open-set condition and the AA, closed-set condition, therefore failing to support hypothesis (3a). Similarly, a Mann-Whitney U revealed a significant difference in slope between the open-set and closed-set conditions, thus failing to support hypothesis (3b). These findings match those of Lay's (2019) sentence lists (Appendices D5 and D6): between the AA, open-set and AA, closed-set, there is a significant difference in SRT ($F(1, 216) = 89.812$, p = <.001, $\eta_p^2 = .294$) and slope (U = 5,088, *p = <.001*).

A visual aid was generated to show the differences and similarities between the open- and closed- set conditions for both the correct UCAMST data and the corrected UCAMST-P data. Separate plots were generated for i) when practice did not immediately precede the condition, ii) when the condition was preceded by practice, and iii) both of these combined.

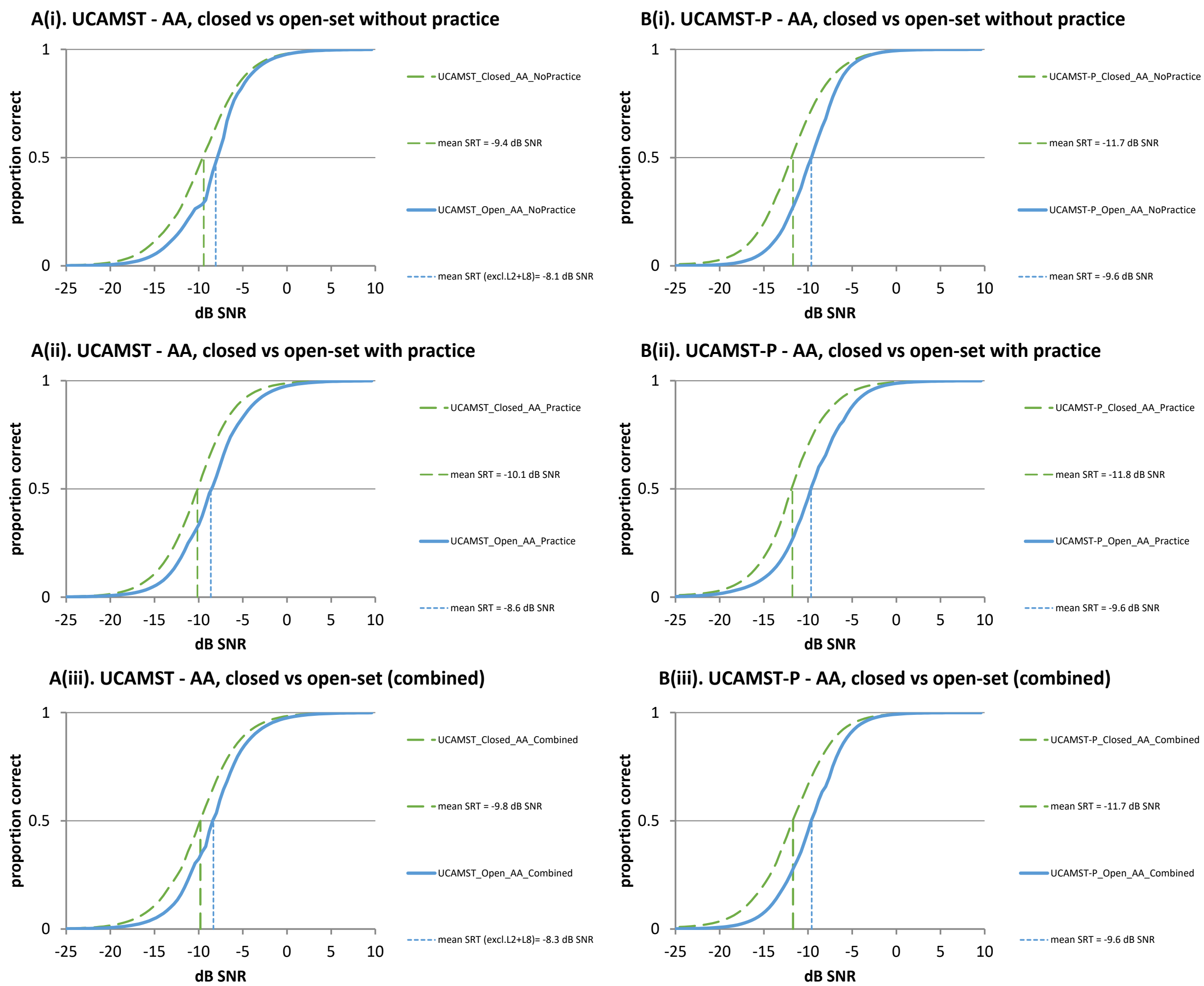


Figure 14. The speech intelligibility functions in the AA mode of presentation for the UCAMST (A) and UCAMST-P (B) both the open- and closed-set response formats generated for i) when practice did not immediately precede the condition, ii) when the condition was preceded by practice, and iii) both of these combined.

PART B

3.5 Response Times

Due to a high number of conditions present, analyses pertaining to response times, hypothesis (4) were made based on graphical data to identify whether there were differences in response times between the UCAMST and UCAMST-P, the high and low SNR, and whether the test was, or was not, preceded by practice. The response time data are displayed in Appendices E1 and E2.

3.5.1 UCAMST and UCAMST-P

Hypothesis (4a) – Response times will be faster for the UCAMST-P than the UCAMST.

There are no apparent differences between the response times for the UCAMST and UCAMST-P under the low SNR condition, regardless of training (Figure 15, (A), (C)). Under the higher SNR condition, the fourth and fifth words of the UCAMST appear slower than the third word of both tests, regardless of training (Figure 15, (B), (D)). Therefore, it appears likely that hypothesis (4a), is supported under high SNR conditions and rejected under low SNR conditions.

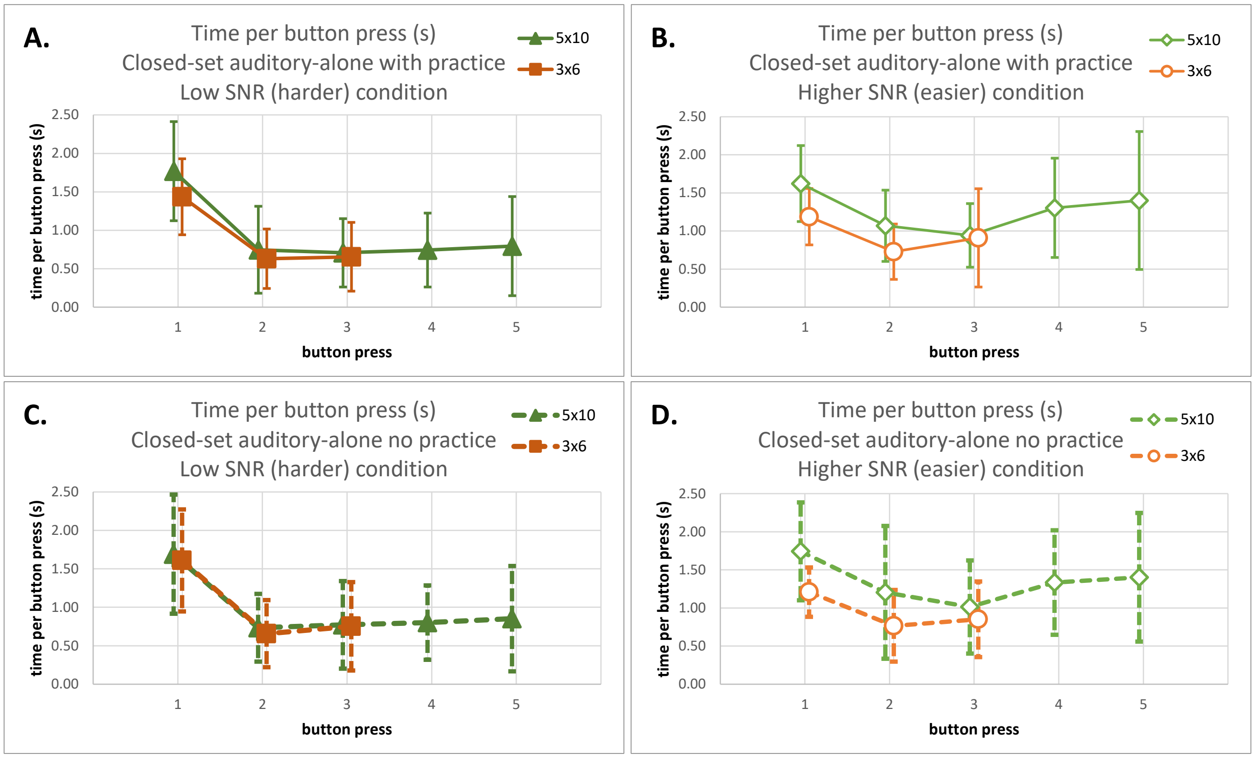


Figure 15. Response times as a function of button press in the AA, closed-set modality for the UCAMST and UCAMST-P, generated for (A) when testing was preceded by practice in the low SNR condition, (B) with practice in the high SNR condition, (C) without practice in the low SNR condition, and (D) without practice in the high SNR condition.

3.5.2 SNR

Hypothesis (4b) – Response times will be faster for the higher SNR than the poorer SNR.

For the UCAMST, the response times for the low SNR stimuli appear faster than the high SNR stimuli, regardless of training status (Figure 16, (A), (C)). There is also no apparent difference between the high SNR and low SNR conditions in the UCAMST-P, regardless of training (Figure 16, (B), (D)). Therefore, it appears likely that hypotheses (4b) is rejected.

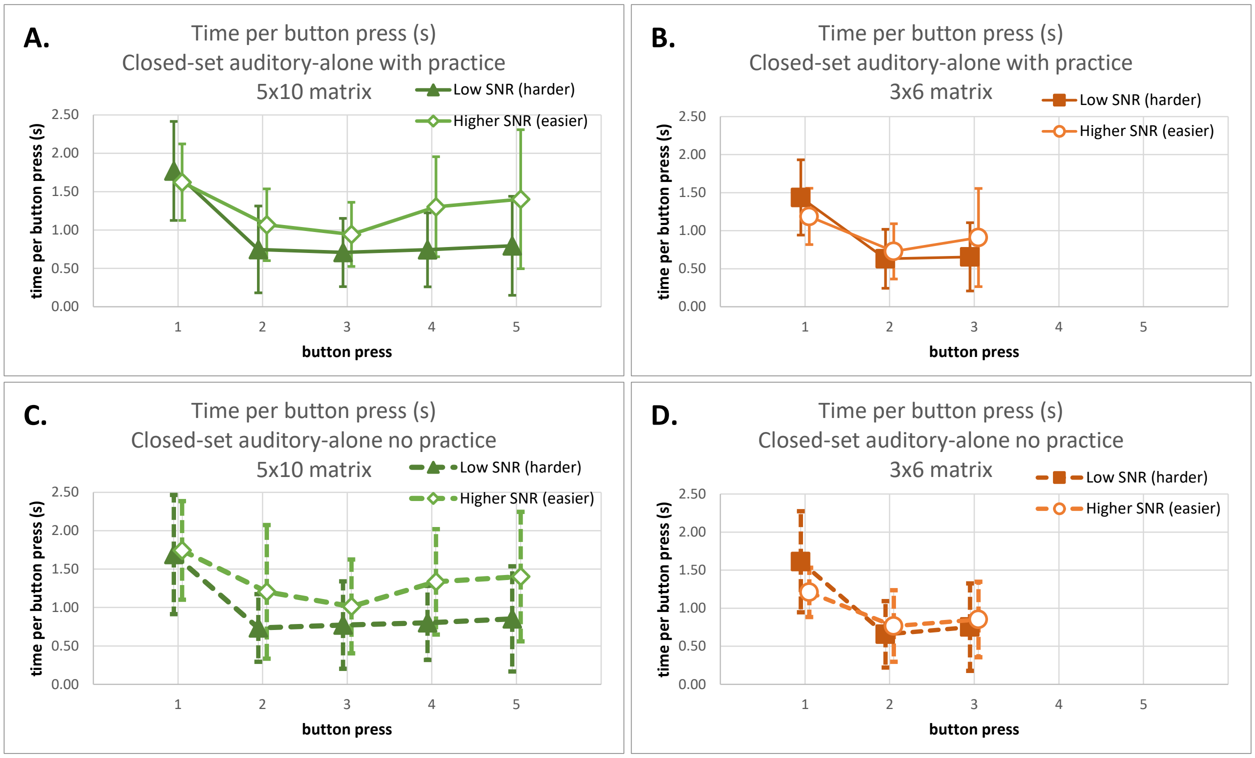


Figure 16. Response times as a function of button press in the AA, closed-set modality for the low SNR and high SNR conditions, generated for (A) when testing was preceded by practice in the 5x10 matrix, (B) with practice in the 3x6 matrix, (C) without practice in the 5x10 matrix, and (D) without practice in the 3x10 matrix.

3.5.3 Effect of practice

Hypothesis (4c) – Response times will be faster for when the test is preceded by practice than, without practice.

No differences appear to exist between when the UCAMST and UCAMST-P matrices are, or are not preceded by training (Figure 17, (A), (B), (C), (D)). Hypothesis (4c) appears to be rejected.

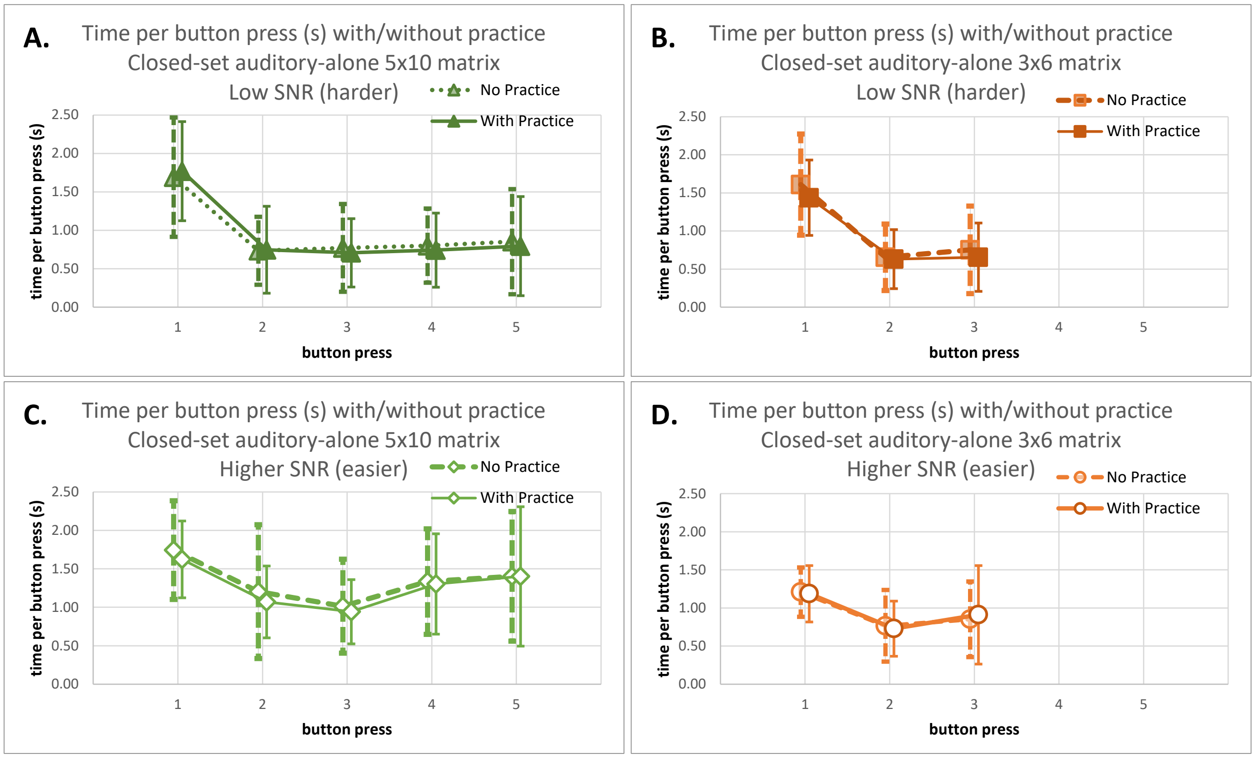


Figure 17. Response times as a function of button press in the AA, closed-set modality comparing the with-practice, and without-practice conditions for (A) the 5x10 matrix in the low SNR condition, (B) the 3x6 matrix in the low SNR condition, (C) the 5x10 matrix in the high SNR condition, and (D) the 3x6 matrix in the high SNR condition.

3.5.4 Response times and accuracy

Response times were also recorded as a function of accuracy under the AA, closed-set condition (Figure 18). To account for the fact that the accuracy is recorded as a proportion correct that is constrained between 0 and 1 (and is therefore not normally distributed), the proportion correct data was subjected to a logit transformation prior to analysis and plotting, such that:

$$\text{logit}(p) = \ln(p/(1-p)).$$

where “p” is proportion correct and “1-p” is proportion incorrect (with p/(1-p) being the odds ratio). Because the logit function is undefined at both p = 0 and p = 1 (due to the inability to divide by zero or take the log of zero), these two proportion correct values were remapped to 0.025 and 0.975 respectively, resulting in logit values that extend from -3.66 to +3.66.

As shown in Figure 18, responses were more accurate for the UCAMST-P compared with the UCAMST under both low and high SNRs, regardless of practice. Note, however, that the low and high SNRs were different for both tests, and were chosen based on predicted (rather than actual) 20% and 80% accuracy points. Responses were more accurate with high SNR conditions than low SNR conditions regardless of whether the test was preceded with practice. There appears to be no effect of practice on accuracy.

As shown in Table 10 below, within each class of data shown in Panels A to D of Figure 18 (e.g. “5x10 low SNR with practice", “3x6 high SNR no practice” etc.), there were poor correlations between the proportion correct and the average response time, with a maximum of 9.25% of variance accounted for by a linear fit for the “5x10 low SNR with practice” condition.

Table 10. Correlation between logit-transformed proportion correct and average time per button press (s) measured within each condition for the UCAMST and UCAMST-P, in the AA, closed-set condition, separated into practice and no practice, and low and high SNR.

Test		Low SNR R^2	High SNR R^2
Practice	UCAMST	0.0925	0.0624
	UCAMST-P	0.0051	0.0796
No Practice	UCAMST	0.0009	0.0414
	UCAMST-P	0.0000	0.0526

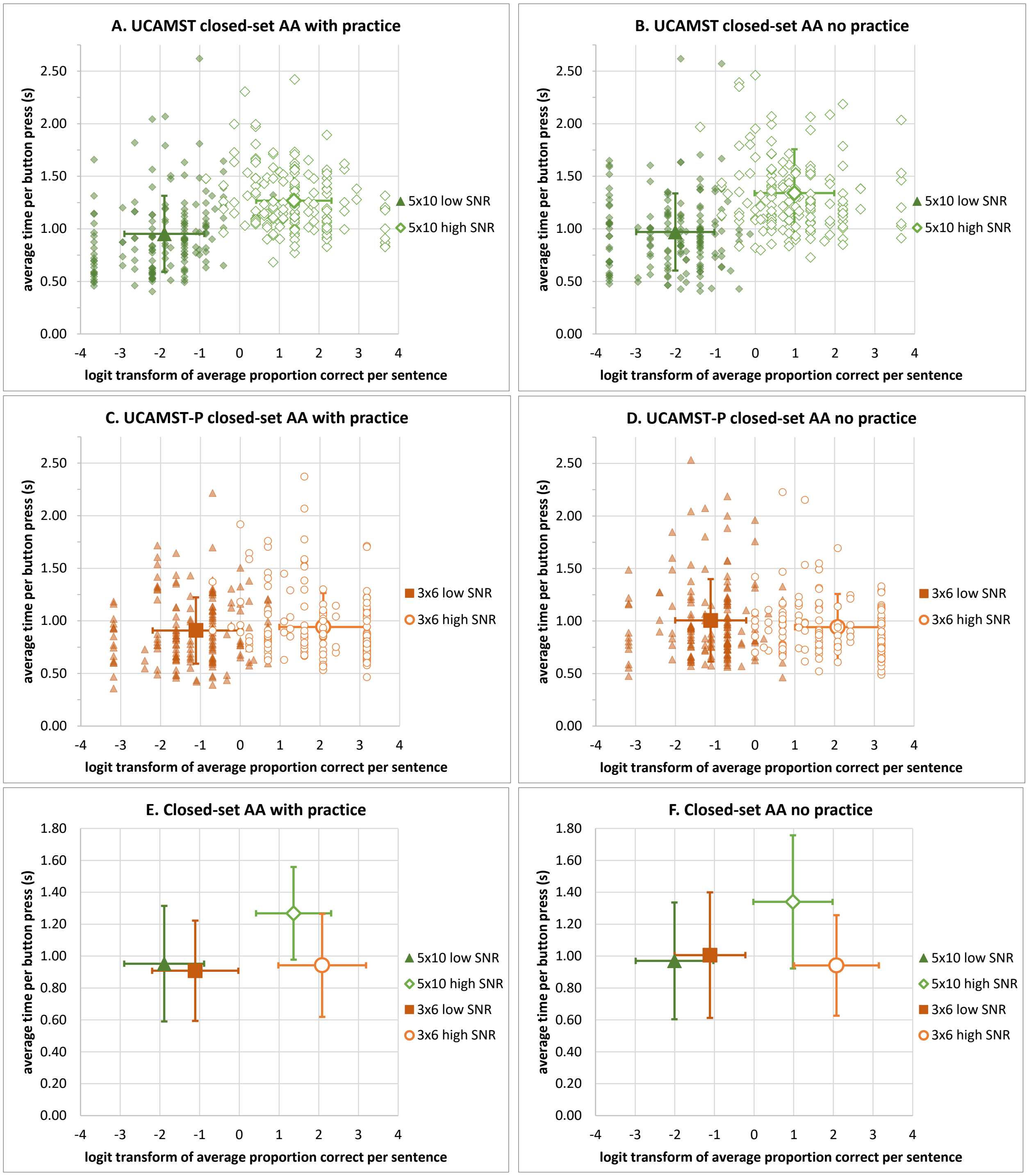


Figure 18. Response times as a function of accuracy in the AA, closed-set modality for low and high SNR conditions, generated for the 5x10 matrix with (A) and without (B) practice, for the 3x10 matrix with (C) and without (D) practice, and to compare the 5x10 with the 3x6 matrices with (E) and without (F) practice.

3.6 Word Selection Order

The results pertaining to research question (4), word selection order, are illustrated in Table 11. A response was defined as "sequential" in the closed-set condition if the response buttons were pushed in the same order that the words appeared in the sentence. Responses were, typically, more sequential for the UCAMST-P compared with the UCAMST (with the exception of the low-SNR no-practice condition), the high SNR compared with the low SNR, and with practice compared with no practice (with the exception of the UCAMST low-SNR condition).

Table 11. Percentage of sequential responses recorded for the UCAMST and UCAMST-P, in the AA, closed-set condition, separated into practice and no practice, and low and high SNR.

Test		Low SNR	High SNR
Practice	UCAMST	77.9%	94.7%
	UCAMST-P	91.5%	98.8%
No Practice	UCAMST	83.9%	88.9%
	UCAMST-P	77.0%	96.3%

Hypothesis (5) – Button pressing will be more sequential for the (a) UCAMST-P than the UCAMST; (b) Higher SNR than poorer SNR; and (c) practice than without practice.

Button pressing was more sequential for the UCAMST-P than the UCAMST with the exception of the low SNR no practice condition, thus partially (i.e. in 3 of 4 conditions) supporting hypothesis (5a). Button pressing was also more sequential for the higher SNR than the lower SNR, regardless of sentence length or practice, thus supporting hypothesis (5b). Button pressing was more sequential for the high SNR conditions, with the exception of the UCAMST low-SNR condition, thus partially (i.e. in 3 of 4 conditions) supporting hypothesis (5c).

Chapter Four: *Discussion*

4.1 Revised UCAMST-P equivalence results

The intent of Part A was to evaluate the UCAMST-P to establish its reliability and sensitivity in estimating SRTs. To do so the SRT values were calculated to evaluate the equivalence of lists under each condition to achieve test-retest reliability. Additionally, the average slope of the speech intelligibility functions were needed to give information about the likely reliability and accuracy of the SRT estimates. Importantly, the SRT and slope values from the AV condition were not comparable to those under the AA condition, with the AV slope values being much lower regardless of which test or response format was used (Jenkins-Foreman, 2018). This effect may be attributed to the advantage offered by lip-reading under the AV condition. Due to the poorer AV data, the AV condition was excluded from this study.

4.1.1 List Equivalence

The first set of hypotheses (1a), (1b), (2a), and (2b) relating to research question (1), proposed that the UCAMST-P sentence lists employed under each test condition (AA, open-set; AA, closed-set) would be equivalent with regards to SRTs and the slopes of the speech intelligibility functions. There were no significant differences between the sentence lists with respect to SRT in the AA, closed-set, nor the AA, open-set condition. These results are an improvement on those originally reported in Jenkins-Foreman (2018), which initially showed significant differences between lists under each condition. However, post hoc testing on the lists performed under the AA open-set condition, did reveal significant differences between lists 2, 8, 12, and 13 with various other lists (Table 5). Consequently, lists 2, 8, 12, and 13 may require further adjustments before the lists can be deemed interchangeable within the AA, open-set condition. Finally, the revised lists generated by Lay (2019) gave similar results as the corrected lists generated by Jenkins-Foreman (2018). In the open-set condition, the mean SRT of the Jenkins-Foreman lists was -9.8 ± 0.9 dB SNR (Appendix C1), compared to -9.8 ± 0.6 dB SNR for the Lay lists (Appendix D1). In the closed-set condition, the mean SRT of the Jenkins-Foreman lists was -11.8 ± 0.4 dB SNR (Appendix C2), compared to -11.8 ± 0.5 dB SNR for the Lay lists (Appendix D2). Taken as a whole, the Lay lists might offer slight advantages in terms of homogeneity of SRT, albeit with a reduced number of lists.

For slope, the results partially align with Jenkins-Foreman (2018); both the erroneous sentence lists (Jenkins-Foreman, 2018) and the new sentence lists (Lay, 2019) were equivalent under AA closed set conditions, and inequivalent under AA open-set conditions. The corrected lists are inequivalent for both response formats. Interestingly, on initial analysis the slope values appeared to have a bimodal distribution. While an attempt was made to filter the data into two groups by slope (slope ≤ 50, and slope >50), insufficient data in the slope >50 group prevented further statistical analysis as separate groups. A greater sample size would give rise to more data and may permit analyses of the two groups separately. This would have the capacity to reveal differences in list equivalency when the slopes are small or large.

4.1.2 Condition Equivalence

The third set of hypotheses pertaining to research question (2), proposed that the open- and closed- set response formats within each test condition (AA, open-set; AA, closed-set) of the UCAMST-P would be equivalent with regards to the slopes of the speech intelligibility functions and SRTs. Equivalence of the test conditions is necessary to establish in order to know whether the conditions may be used interchangeably.

There was a significant difference between the open- and closed- set response formats with regards to SRT in the AA modality (AA open-set vs AA closed set) in the lists of both

Jenkins-Foreman (2019) and Lay (2019). This was unsurprising given that Jenkins-Foreman (2018) had previously shown significant differences in SRT between the open- and closed- set response formats for both the UCAMST and the UCAMST-P, within each modality (AA open-set vs AA closed set; and AV open-set vs AV closed-set). Previously Jenkins-Foreman (2018) has established higher SRTs in the open-set condition for AA and AV presentation modes for both the UCAMST and the UCAMST-P. This finding was repeated in the current study; SRTs were higher in the open condition (AA, open-set) than the closed (AA, closed-set). This finding enforces that the two conditions cannot be used interchangeably.

Listeners with a HI have also achieved higher SRTs on the UCAMST in AV open-set, compared with AV closed-set conditions, signifying that the open-set is harder in the AV modality (Andre, 2016). Participants in Andre's (2016) study of the 5-word UCAMST were over 60 with a HI so estimates may have been confounded by WMC (Andre, 2016; Theunissen et al., 2009). However, varying estimates of SRT are given by other literature under the open- and closed- set formats. SRTs have been shown to be significantly higher in the closed-set condition, suggesting that the closed-set condition is more difficult (Hochmuth et al., 2012; Stone, 2016). This may be attributed to greater cognitive demands necessitated in closed-set as listeners have to hold onto the sentence while they select the words in the matrix. Ozimek and colleagues (2010), on the other hand, found no differences between the response formats for the Polish MST. As the literature regarding the effect of response format on SRT is conflicting, it may be best to decide the response format for the UCAMST and UCAMST-P based on the ability and needs of the individual.

No significant differences were found between the two conditions with regards to slope, aligning with previous research (Jenkins-Foreman, 2018). Consequently, while the accuracy of estimating the SRT is equivalent regardless of using open or closed set response formats, the SRT estimates themselves may differ depending on the response format and modality.

4.2 Working Memory Capacity

4.2.1 UCAMST and UCAMST-P

Regardless of the UCAMST-P being fundamentally shorter, under high SNR conditions, the response times for the 3-word pseudo-sentences remain faster than for the first three words presented in UCAMST. Notably, recalling three words and locating them on the 3x6 matrix is likely less cognitively demanding than recalling five words and locating them on the 5x10 matrix, which has over twice as many words.

Conversely, there appear to be no significant differences between the response times of the two tests under low SNR conditions. As discussed previously (section 1.6.3), a signal presented with no masking noise is understood automatically and effortlessly through an implicit process. When this signal is compromised, for example with a poor SNR, an explicit process that exhausts WMC is necessary to make use of the compromised signal (Rönnberg et al., 2013). This process requires more time than is necessary for listening in uncompromised environments with a high SNR. As a result, the cognitive demands of the low SNR condition may exceed the difference in cognitive demands between the sentence length of the two tests. Furthermore, the response times of the fourth and fifth chosen words of the UCAMST under the high SNR condition appear slower than the third, suggesting that participants may chunk the first three words in their memory and have to engage cognitive resources to recall the final two. This provides additional evidence that the length of the three-word pseudo-sentence is less cognitively demanding and therefore, more appropriate for populations with decreased WMC.

4.2.2 High SNR and Low SNR

As discussed in section 2.1.3, the signals for the closed-set response format were presented alongside constant masking noise at two different SNRs: -14.0 dB SNR and -7.4 dB SNR. These were chosen to correspond to the pair of compromise, 80% and 20% correct in order to simultaneously estimate the slope and SRT (Brand & Kollmeier, 2002).

Hypothesis (4b), relating to research question (3), proposed that response times would be faster for the higher SNR sentences compared with the lower SNR sentences owing to the fact that there is a greater cognitive load. Interestingly, the inverse was true for the UCAMST; faster reaction times were recorded for the lower SNR compared with the higher. This may be attributed to the psychological effect of rapid guessing behaviour which occurs in low-stakes testing. This is where a listener does not believe they know the answer and responds with such speed that they could not have had enough time to actively read and consider their response (Schnipke & Scrams, 2002). As a result, accuracy is at or near random. This is compared with solution behaviour, in which an examinee actively and carefully seeks to answer each item correctly (Mills et al., 2014; Schnipke, 1995). Consequently, the accuracy of the responses in low-stakes testing is dependent upon the difficulty of the item or other item characteristics, as well as the examinees ability (Mills et al., 2014). This is shown in the findings pertaining to the UCAMST as the lower SNR yielded both poorer accuracy, and faster responses than the high SNR condition (Figure 18).

This effect may have been present in similar research by Meister and colleagues (2018) who illustrated that reaction times decreased for intelligibility levels below 50 percent. Rapid guessing behaviour may then be evident in speech recognition testing when the listening conditions are too difficult. In the current study, for example, there were faster response times for the 5-word sentences under low SNR (harder) conditions compared with those of the 5-word sentences under the high SNR (easier) conditions. In an attempt to combat the effect that rapid response behaviour had on test validity, Wise and Kong (2005) developed an index of examinee effort based on item response times in computer-based tests, believing that rapid responses are due to poor participant effort. Following rapid-response filtering and rescoring answers that were attributed to rapid guessing to missing, the validity of the test score was increased. Rapid-response filtering could be a feasible option to achieve test scores with improved validity.

On the other hand, the UCAMST-P had no difference between the two different SNR conditions. In low-stakes testing, Wise (2006) found that the strongest predictors of how much effort each item received were item length (amount of necessary reading or scanning required) and item position. Items of higher difficulty that place greater load on participant's working memory, are also more likely to be guessed under low-stakes conditions (Bovaird, 2002). Evidently, the five-word sentences of the UCAMST, were at greater risk of being guessed compared with the three-word pseudo-sentences. In this way, the three-word pseudo-sentences were perhaps more manageable for the participant's WMC, and may provide some evidence of the suitability of the UCAMST-P for populations with decreased WMC.

4.2.3 Word Selection Order

With the exception of the UCAMST under low SNR conditions, both MSTs yielded a slightly greater proportion of sequential results following practice than without practice. However, the significance of these findings were not calculated and thus must be applied with caution. It is possible that the two training lists allowed participants to ease into the test and develop a technique where it may be more efficient to answer sequentially. Notably however, specific instructions were not provided to participants about whether they were required to respond sequentially.

Generally, responses were also more sequential under the easier (high SNR, 3-word) conditions compared with harder (low SNR, 5-word) conditions. This was expected as the lower SNR is more difficult to hear, and the 5-word sentence requires more words be recalled, so participants may recall words that they were more certain they heard correctly, prior to words they are less sure about. Previous research conducted by Murdock and Okada (1970) on free-recall discusses that the duration between each recalled item, or inter-response times (IRTs), are short when neighbouring list items are recalled successively. However, the IRT increases as the distance between the item's positions in the study list increases, drawing an association between greater processing time and non-sequential responses. This shows that non-sequential responses are probably due to the test exceeding the individual's WMC.

4.2.4 The Training Effect

The training effect is the amount of training prior to a speech recognition tests that is necessary to stabilise the SRT measurements (Dietz et al., 2014; Hagerman, 1984; Hochmuth et al., 2012; Kollmeier et al., 2015; Wagener et al., 2003). Notably, an individual's SRT levels are shown to decrease as more lists are administered (Wagener et al., 2003). Significant training effects have been recorded for various MSTs internationally, particularly for first time users within the first few lists (Dietz et al., 2014; Hochmuth et al., 2012; Kollmeier et al., 2015; Puglisi et al., 2015; Wagener et al., 2003; Warzybok et al., 2015). In the current study, training was implemented before each block condition (section 2.1.5) to let participants learn the task before starting.

Whether or not the 5-word test was preceded by training did not cause any apparent difference on response times or response order. This aligns with previous results from Jenkins-Foreman (2018) who found that for the UCAMST, training does not impact the SRT or the slope in the AV presentation mode, nor does it impact the slope in the AA presentation mode. However, a significant difference in SRT was previously established between the training and no training conditions of the UCAMST under the AA presentation mode. Similarly, whether or not the 3-word test was preceded by training had no notable differences on response times, aligning with Jenkins-Foreman's (2018) findings that training does not impact the SRT, nor the slope, in any of the presentation modes (AA, open-set; AA, closed-set; AV, open-set; AV, closed-set) for the UCAMST-P. While it appears that the NZ MSTs are not significantly affected by the training effect thus far, removing the practice lists of the UCAMST-P, in particular, may be premature. Training on cognitive tasks has been shown to assist both older populations (Glisky, 2007), and children (Rowe et al., 2019) to adjust to cognitive tests and thus improve performance, and this is reflected in the finding of other international paediatric MSTs.

In the German Paediatric MST (OlKiSA; Neumann et al., 2012), a training effect was identified for NH young adults. Like the 5-word MSTs, improvements were greatest in the first two test measurements (1.5 dB SNR), followed by no significant differences between the test lists after the second set measurements (Buschermöhle et al., 2016). Similarly, clear improvements in the SRT estimates were identified during the training phase of the Finnish Paediatric MST (FINSIMAT; Willberg et al., 2020). Consequently, Willberg and colleagues (2020) stated that familiarising the participants with the test material in written form and conducting two training measurements prior to the actual measurement seems sufficient to eliminate clinically relevant learning-related improvements during a test session. It has been recommended that regardless of language, two practice lists of 20 sentences each, should be given before the MST to help the SRT measurements to stabilise (Dietz et al., 2014; Hochmuth et al., 2012; Kollmeier et al., 2015; Wagener et al., 2003). Although, Willberg and colleagues (2020) also suggest that as the inclusion of sentence lists increases the duration of the test, it may increase the test's cognitive demands and increase the individual's apparent SRT. As a result, further testing of both the UCAMST and UCAMST-P is required within their target populations before practice lists can be deemed unnecessary.

4.3 Study Limitations

The presence of certain limitations in the current study may challenge the results and must be considered when interpreting and applying the findings. The following section examines these limitations and discusses how further research may remedy these shortcomings.

4.3.1 The Sample

4.3.1.1 Lack of Normality

Non-parametric tests were employed on data pertaining to the slope for the UCAMST-P due to significant bias in the data, violating the assumption of normality. Subsequently, post-hoc analyses also could not be run. Therefore, sentence lists that significantly differed from other lists with respect to slope, could not be identified.

4.3.1.2 Generalisability

The purpose of the current study was to evaluate list and condition equivalency under the AA modality for the UCAMST-P, and to evaluate the impact of auditory memory on the UCAMST and UCAMST-P. Both of these aims were tested on a sample of participants who likely represent the performance typically expected for NH listeners. However, the data may not truly represent the actual age, linguistic, cultural, and SES variations in NZ due to how people were recruited. The sample had a large proportion of students from the University of Canterbury (Christchurch, New Zealand), and 85% of the sample were between the ages of 20-30-years. Furthermore, far more females (n=32) than males (n=11) participated in this research. It is unclear whether this gender imbalance is notable, as the evaluation of previous MSTs have revealed discrepancies between the performance of males and females, based on differences in speech reading abilities (Ozimek et al., 2012; Stone 2016; Wagener et al., 2003).

Furthermore, previous literature suggests that WMC can be influenced by age, sex, and education level (Cansino et al., 2013; Pliatsikas et al., 2018). For example, with increasing age, males have shown a steeper decline in WMC compared with females, and with increasing education, females had greater gains in WMC compared with males (Pliatsikas et al., 2018). Cansino and colleagues (2013) found that males between the ages of 41 and 50 performed better on auditory working memory tasks than females of the same age bracket, and performed better than females across a mixture of general WMC tasks between the ages of 21 and 30, and 41 and 60. This study cannot conclude that the results can be generalised for paediatric, and older populations, particularly as WMC in this population is reduced (van Rooij & Plomp, 1990). A more representative sample should be employed in future research to improve the generalisability of further findings.

4.3.1.3 Touch Screen Interface and Age

When employing a touch screen, consideration must be given to whether some individuals may have greater difficulty localising their response on a screen, regardless of memory span. The literature on usability of touch screen interfaces is sparse. Although, it is well documented that age-related decline in motor ability directly interferes with an older person's ability to use touch screen devices due to reduced capacity for: fine motor control (Holmes et al., 2009), a propensity for making more sub movements and being slower to capture a target using a mouse (Smith et al., 1999; Walker et al., 1996), and having a general slowing of movement, particularly with repetitive-speed tasks (Holmes et al., 2009; Hawthorn, 2000). Gao and Sun (2015) found that older participants (52-81 years-old) were slower at locating targets on a touch screen than younger participants (19-24 years-old). Even in older individuals with normal or near-normal vision, approximately half were found to have difficulty locating objects of interest in the environment (Owsley et al., 1995). Conversely, Murata and Iwase (2005) found no difference in pointing time and target location on a touch screen interface between young, middle aged, and older adults. Nonetheless, based on motor ability, it is expected that older populations will have longer response times regardless of the auditory memory capacity. This raises the importance with which age-related normative data is collected.

4.3.2 Statistical Analyses

No formal statistical analyses were employed in the analyses of the response time and response order data (Part B) due to time constraints. Consequently, it is not clear of the effect sizes, nor significance, or results pertaining to WMC. While it appears clear that the UCAMST-P is less cognitively demanding than the UCAMST, the statistical significance of this finding is unclear.

4.3.3 Baseline Reaction Data

As the data was obtained as a subset of a larger data set from Jenkins-Foreman (2018), changes could not be made to the way data was collected. Measuring the participants baseline reaction time and recording a measure of their WMC could have been beneficial prior to initiating testing with the UCAMST and UCAMST-P, to account for differences in each individuals response times and reduce variance attributed to individual factors. These points may be important for future research, particularly on paediatric and senior populations where there is likely to be more variance in response times and WMC.

4.3.5 Absent Conditions

4.3.5.1 Absence of open-set condition for assessing auditory memory

Due to the nature of response times being recorded with a button press in the open-set format, reaction times were simply recorded. On the other hand, open set responses required verbal repetition and reaction times were based on how quickly the assessors entered the participant's response (Jenkins-Foreman, 2018). While not analysed in this study, it is possible that a difference may be found between a participants performance in the open and closed sets. In the open-set, for example, there is no visible matrix with which to search for words as the listener verbally repeats the sentence. This may decrease a participant's cognitive load and decrease reaction times. However, this may also be more cognitively demanding on the listener as they are not reassured when selecting an answer on the matrix. Consequently, the findings pertaining to WMC in the present study cannot be applied under open-set conditions.

4.4 Future Research

4.4.1 Piloting with different demographics

To date, the research preparing the UCAMST and UCAMST-P for clinical application in NZ has been primarily focused upon young adults with NH (Jenkins-Foreman, 2018; McClelland, 2015; Ripberger, 2018; Stone, 2016). For the initial stages of testing this demographic is sufficient, however, the UCAMST and UCAMST-P must be tested on a substantially broader demographic in order to be applicable to further groups in New Zealand. In order to gain normative data, and optimize speech recognition tests, high equivalence test stimuli and word-specific intelligibility functions are required for each word. This is generally achieved by presenting speech material to individual's with NH at fixed SNRs, and is followed by the exclusion of materials that, following adjustments, fail to fit the word-specific intelligibility function (Akeroyd et al., 2015)

4.4.1.1 Paediatric populations

Trialling the UCAMST-P with children is vital considering its intended purpose is for assessing paediatric populations. As working memory gradually improves through childhood (Brocki & Bohlin, 2004; Gathercole et al., 2004; Lendinez et al., 2015; Luna et al., 2004) and a child's ability to understand speech in noise develops with age (Buss et al., 2019; Corbin et al., 2016; Stuart, 2005), it is unsurprising that there are clear differences in the reliability of results between adult and paediatric populations. For example, school-aged children had less reliable results than adults on the German MST (OLSa; Wagener et al., 1999a-c) which was attributed to childrens' poorer auditory memory capacity (Wagener & Kollmeier, 2005). This is in accordance with further literature that shows an increased variability of scores with decreasing age due to greater developmental variability in younger age groups (Holder et al., 2016; Ng et al., 2011; Wilson et al., 2010).

Lay (2019) piloted the UCAMST-P with children between 6-12-years-old with NH, and measured the effect of age, among other factors, on UCAMST-P performance. Testing was performed in the AA, open-set condition, to avoid a reading artefact that would likely arise due to the range of reading abilities in younger age brackets (Holder et al., 2016). SRT scores improved with age until approximately 10 years, then plateaued for the remaining age groups (Lay, 2019). Due to a small sample size (20 participants per age group), Lay (2019) commented that further investigation with more participants my improve the study's accuracy; this was a small population with which to base normative data, making differentiating outliers that could skew the results difficult. Additionally, participants were recruited from Christchurch schools which may not necessarily be representative of New Zealand due to socioeconomic, ethnic, geographical, cultural differences. It may benefit to introduce testing in an additional location or undertake sampling across NZ to add value to this normative data.

Several additional points would add value and should be considered when initiating further collection of normative data. First, Lay (2019) looked solely at the AA, open-set condition. Further research should look into collecting normative data when employing both AV, and closed-set conditions in addition to AV, and open-set conditions. Second, extending the age limit of the study to incorporate younger populations (i.e. 3 to 5-year-olds) would test the viability of the UCAMST-P in this population and may provide guidance on the developmental trajectory of SRT scores as young children mature. Finally, as performance in speech recognition measures can be influenced by whether reinforcement is employed in testing (Kirk et al., 1997), Kosky and Boothroyd (2003) suggest that paediatric speech recognition measures should be interesting and motivating, in order to obtain a more accurate SRT.

As the length of paediatric speech recognition test is defined by the child's fatigue (Kosky & Boothroyd, 2003; Neumann et al., 2012), motivating factors should be considered. For example, visual or auditory reinforces presented between presentations may improve the child's attention and motivation to complete the test. The UCAMST-P may also benefit from a 67% correct threshold compared with the typical 40% correct threshold, in order to maintain optimism and motivation in younger populations (Hagerman & Hermansson, 2015). It is necessary, therefore, to establish normative data to ascertain which age-range the UCAMST-P is appropriate for and to validate its outputs (Neumann et al., 2012), as well as introducing features to make the task more motivating and thus potentially more accurate for children.

4.4.1.2 Older populations

Adults older than 65-years-old make up the bulk of those attending audiology clinics so accurate and efficient measures of speech recognition are crucial (Newman & Sandridge, 2004). There is evidence that one's performance in speech in noise recognition measures declines with age (Humes, 2015) and may be attributed to WMC (Arehart et al., 2013; Arlinger et al., 2009; Foo et al., 2007; Rudner et al., 2012). Further validation research should therefore be conducted on elderly listeners, particularly those with a cognitive impairment or dementia, using both the UCAMST and UCAMST-P to gain normative data in order to validate the test's results. This is particularly important as while a shorter test may appear more suitable for individuals with limited WMC, there is some risk that less test items may reduce the reliability and repeatability of the results (Willberg et al., 2020).

4.4.1.3 Individuals with a HI

As it currently stands, the UCAMST and UCAMST-P have only been tested on populations with NH. However, a speech in noise test must be able to evaluate an individual's speech recognition capacity regardless of whether the person has NH or a HI. As such, the UCAMST and UCAMST-P must be tested on populations with varying degree of HI in order to validate their use in these settings. As previously discussed, those individuals with a HI struggle hearing in noise more than normal hearing individuals (Lewis et al., 2016; Ng et al., 2011; Peters et al.,1998; Wilson et al., 2007). In clinical settings, this causes greater variance between the expected SRT in individuals with a HI compared with individuals with NH. (Peters et al., 1998; Wilson et al., 2007).

Notably, the difference between individuals with NH and those with a HI is also exacerbated when using babble noise which is attributed to the effect of masking-release. This is where dips in the acoustic masker allow a listener to 'glimpse' the target signal. However, the effect is typically negligible for those with a HI which further divides the SRT's between individuals with NH and with a HI (Bacon et al., 1998; Hopkins & Moore, 2009). It is pertinent then that the expected performance for individuals with a HI is calculated using multi-talker (babble) noise, speech-shaped (constant) noise, and in quiet (Wagener & Kollmeier, 2005). This would give rise to normative data that individual's performances can be compared with to ascertain how well they can understand speech in noise (Akeroyd et al., 2015).

Preparing the UCAMST and UCAMST-P for clinical application now requires normative data under each presentation mode and modality for listeners of different ages and with different degrees of HI. This will provide an index with which an individual's performance can be compared with, to find a relative measure of their ability to understand speech in noise.

4.4.2 Trialling a Picture-Pointing Response

Administering speech recognition tasks to children requires careful selection of the response method as open- and closed-set response formats both have limitations. While open-set formats evaluate lexical memory and acoustic-phonetic activation, difficulties arise when listeners have limited language or unusual speech pronunciation, due to a HI, for example (Calandruccio et al., 2014; Ozimek et al., 2012). On the other hand, closed-set requires that an individual to respond based on phonological and lexical competition (Clopper et al., 2006). As closed-set requires selection of words on a visible matrix, it is susceptible to guessing bias and training effects (Ozimek et al., 2012). Furthermore, in children who are particularly young, or have limited literacy, selecting a word on a screen is not always feasible (Calandruccio et al., 2014). Consequently, using a picture-pointing, closed-set measure is recommended as a behavioural assessment tool as it can effectively evaluate speech recognition in paediatric populations (Hall et al., 2002; Litovsky, 2005; Ross & Lerman, 1970). Currently a picture pointing response format is available in the PPMST (designed for children between 3-6-years-old), where each word in the matrix corresponds with an image (Ozimek et al., 2012). The images are arranged in a six-picture matrix containing the target word, and associated alternatives. A picture-pointing response method option would be a valuable addition to the UCAMST-P so that performance is not impacted by speech production or limited literacy levels, particularly as test is intended for these paediatric populations (Kosky & Boothroyd, 2003)

4.4.3 Analysis of AV Modality

Data from Jenkins-Foreman (2018) also permits comparison of the response times between AA and AV conditions. Granting a listener access to both auditory and visual speech signals provides complimentary information that is effectively combined so that speech understanding under AV conditions exceeds that of speech understanding in AA conditions (Sommers et al., 2005; Tye-Murray, et al., 2007). Pillai and Yathiraj (2017) compared AA, AV, and VA (visual-alone) modalities across several different measures of memory in children and found no significant difference between the AA and AV modalities on memory. Furthermore, presenting a list of free recall auditorily results in a larger recency effect (better memory for the last few list items) than presenting the list visually (a modality effect). However, these tests did not employ background noise and used different stimuli. From here, identifying whether there is a difference between the AA and AV modalities with respect to response times of the UCAMST and UCAMST-P is necessary whether the modalities can be used interchangeably, or if one is more appropriate than the other for populations with reduced WMC.

4.4.4 Cross-validation with other speech tests

4.4.3.1 UCAMST

Andre (2016) established a correlation between the results obtained from the UCAMST and the Quick Speech in Noise Test (QuickSIN), and Ripberger (2018) has cross-validated the UCAMST with the meaningful CVC words recognition test (Boothroyd, 1968; Boothroyd & Nittrouer, 1988; Purdy et al., 2000) and the QuickSIN (Killion et al., 2004). This research has shown that the UCAMST has the potential to replace the meaningful CVC word recognition test in practice (Ripberger 2018). However, the UCAMST also bears significant similarities to the New Zealand Hearing in Noise Test (NZHINT; Hope, 2010). This test uses NZ English sentences as stimuli and is presented with constant masking noise. It also has similar phonemic list balancing and sentence intelligibility equalisation, while using adaptive procedures to locate a listeners SRT (Hope, 2010; McClelland, 2014). Notably, the NZHINT is unreleased, not normalised, and not validated. It would take significant work to move it toward commercial release. Nonetheless, it may be valuable to examine the relationship between the two tests to examine whether they may be used interchangeably. Cross validating the UCAMST with the NZHINT may then be an area for future research to identify decreasing redundancy across speech in noise measures available in NZ clinical settings.

4.4.3.2 UCAMST-P

Future research should cross validate the UCAMST-P with common speech recognition measures employed in the NZ test battery. This could assess which tests give information complementary to that provided by the UCAMST-P. As the KTT is the common test of choice for paediatric speech recognition testing in NZ clinical practice, cross-validation of the UCAMST-P with the KTT is warranted (Antognelli, 1986). Similarly, validating the

UCAMST-P with the QuickSIN may be beneficial as it also uses masked sentence stimuli (Killion et al., 2004). Comparisons between the UCAMST-P and other commercially available speech recognition tests give better awareness of what information can be gained from the UCAMST-P compared with other available measures, for example, which tests are the most time efficient and best for inclusion in a paediatric test battery (Wilson et al., 2007).

4.4.5 Confusion Matrices

Further data collected by Jenkins-Foreman (2018) includes several confusion matrices for each test. Under the closed-set condition, it can be seen when individual words in a sentence are incorrectly swapped, and for which word they have been confused with. For example, when sentence 41 of the UCAMST was presented in the closed-set format at -14 dB SNR (-13.97), the word “twelve” was only recognised correctly 33% of the time. Otherwise, it was incorrectly recognised as “nine” 33% of the time, and “ten” 33% of the time. Similarly under the open-set response format, the matrices recorded the proportion of words that were answered correctly or incorrectly for each sentence of the UCAMST and UCAMST-P. For example, using sentence 41 in the AA condition of the UCAMST in an open-set response format (at -11.62 dB SNR), the word “Kathy” was correctly recognised 25% of the time and recognised incorrectly 75% of the time. These confusion matrices may give insight into which words are frequently answered incorrectly.

4.4.5 Transducer

Research on the UCAMST and UCAMST-P has been conducted using supra-aural headphones to present the stimuli. In some clinical settings, speech recognition tests are conducted in sound-field settings so that the listener can be tested with their amplification device to objectively verify a hearing aid fitting, or in cases where children do not tolerate headphones.

However, sound-field presentation can cause sound reflection off surfaces in the test area which may degrade speech intelligibility and subsequently influence SRTs, particularly if the test is not conducted in a sound-treated environment (Allen & Berkeley, 1979; Soli & Wong, 2008). Further research could be conducted to assess whether the UCAMST and UCAMST-P are appropriate for sound-field use.

4.5 Concluding Statements

Speech audiometry is a crucial component of the audiological test battery as it provides valuable information about an individual's capacity to communicate in everyday life. However, current tests of speech recognition used in NZ clinical practice present with various limitations pertaining to their use of single-words and quiet test settings, which do not reflect typical daily communication. MSTs, on the other hand, are shown to assess a combination of signal processing, background noise, and individual WMC more consistently than with standard HINT testing (Plomp-type sentences) (Rudner et al., 2009; Rudner et al., 2011).

Due to the potential cognitive demands of the UCAMST, the UCAMST-P was created as a simplified speech in noise test for native New Zealand speakers. It was specifically produced to extend the current speech test battery for populations whom have difficulty with the complexity of the currently available speech in noise tests. Following the amendment of data from Jenkins-Foreman (2018), the current research showed that the UCAMST-P sentence lists are equivalent under the AA, open-set and AA, closed-set conditions. However, as was also shown by previous research, these UCAMST-P conditions will require their own condition-specific normative data. From here, piloting the UCAMST-P with its intended populations and recording normative data must be achieved before the test can be implemented into practice.

Finally, the current research on the response times and response orders of the UCAMST-P suggests that it is less cognitively demanding and more suitable than the UCAMST for those with reduced WMC, including paediatric and older populations. These findings are valuable in the ongoing production of the UCAMST and UCAMST-P, and highlight areas where further adjustments and analyses are required before the integration of the tests into NZ clinical practice.

References

Akeroyd, M. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal Of Audiology*, *47*(sup2), S53-S71. https://doi.org/10.1080/14992020802301142

Akeroyd, M., Arlinger, S., Bentler, R., Boothroyd, A., Dillier, N., & Dreschler, W. et al. (2015). International Collegium of Rehabilitative Audiology (ICRA) recommendations for the construction of multilingual speech tests. *International Journal Of Audiology*, *54*(sup2), 17-22. https://doi.org/10.3109/14992027.2015.1030513

Alhanbali, S., Dawes, P., Lloyd, S., & Munro, K. (2017). Self-reported listening-related effort and fatigue in hearing-impaired adults. *Ear And Hearing*, *38*(1), e39-e48. https://doi.org/10.1097/aud.0000000000000361

Allen, J., & Berkley, D. (1979). Image method for efficiently simulating small-room acoustics. *The Journal Of The Acoustical Society Of America*, *65*(4), 943-950. https://doi.org/10.1121/1.382599

Alloway, T., Gathercole, S., & Pickering, S. (2006). Verbal and visuospatial short-term and working memory in children: are they separable?. *Child Development*, *77*(6), 1698-1716. https://doi.org/10.1111/j.1467-8624.2006.00968.x

Antognelli, T. (1986). *The Kendall Toy Test Revisited: The Development of an Australian Version.* Conference presentation, VII National Conference of the Audiology Society of Australia.

Arehart, K., Souza, P., Baca, R., & Kates, J. (2013). Working memory, age, and hearing loss. *Ear And Hearing*, *34*(3), 251-260. https://doi.org/10.1097/aud.0b013e318271aa5e

Arlinger, S., Lunner, T., Lyxell, B., & Pichora-Fuller, M. (2009). The emergence of Cognitive Hearing Science. *Scandinavian Journal Of Psychology*, *50*(5), 371-384. https://doi.org/10.1111/j.1467-9450.2009.00753.x

Aronen, E., Vuontela, V., Steenari, M., Salmi, J., & Carlson, S. (2005). Working memory, psychiatric symptoms, and academic performance at school. *Neurobiology Of Learning And Memory*, *83*(1), 33-42. https://doi.org/10.1016/j.nlm.2004.06.010

Bacon, S., Opie, J., & Montoya, D. (1998). The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds. *Journal Of Speech, Language, And Hearing Research*, *41*(3), 549-563. https://doi.org/10.1044/jslhr.4103.549

Baddeley, A. (1986). *Oxford Psychology Series, No.11. Working memory*. Clarendon Press/Oxford University Press.

Baddeley, A., & Hitch, G. (1974). Working Memory. *Psychology of learning and motivation*, *2*, 47-89. https://doi.org/10.1016/s0079-7421(08)60452-1

Baddeley, A., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal Of Verbal Learning And Verbal Behavior*, *14*(6), 575-589. https://doi.org/10.1016/s0022-5371(75)80045-4

Beattie, R., Barr, T., & Roup, C. (1997). Normal and hearing-impaired word recognition scores for monosyllabic words in quiet and noise. *British Journal Of Audiology*, *31*(3), 153-164. https://doi.org/10.3109/03005364000000018

Beechey, T., Buchholz, J., & Keidser, G. (2020). Hearing impairment increases communication effort during conversations in noise. *Journal Of Speech, Language, And Hearing Research*, *63*(1), 305-320. https://doi.org/10.1044/2019_jslhr-19-00201

Bell, T. S., & Wilson, R. H. (2001). Sentence recognition materials based on frequency of word use and lexical confusability. *Journal of the American Academy of Audiology, 12*(10), 514-522.

Benson, E., McCreery, R., & Jorgensen, L. (2018). Conventional amplification for children and adults with severe-to-profound hearing loss. *Seminars In Hearing*, *39*(04), 364-376. https://doi.org/10.1055/s-0038-1670699

Billings, C., & Madsen, B. (2018). A perspective on brain-behavior relationships and effects of age and hearing using speech-in-noise stimuli. *Hearing Research*, *369*, 90-102. https://doi.org/10.1016/j.heares.2018.03.024

Boboshko M, Warzybok A, Zokoll MA, Maltseva N (2013) RUMatrix test: construction, evaluation and clinical validation. *Otorhinolaryngologia Hungarica*, Vol. 59, N 2.-P. 49 (Abstracts of the 11th EFAS Congress Budapest).

Boothroyd, A. (1968). Developments in speech audiometry. *British Journal Of Audiology*, *2*(1), 3-10. https://doi.org/10.3109/00381796809075436

Boothroyd, A. (2008). The performance/intensity fFunction: An underused resource. *Ear And Hearing*, *29*(4), 479-491. https://doi.org/10.1097/aud.0b013e318174f067

Boothroyd, A., & Nittrouer, S. (1988). Mathematical treatment of context effects in phoneme and word recognition. *The Journal Of The Acoustical Society Of America*, *84*(1), 101-114. https://doi.org/10.1121/1.396976

Buschermöhle, M., K. C. Wagener, and B. Kollmeier. (2016). "Sprachaudiometrische Messungen mit dem Verkürzten Oldenburger Satztest OLKISA bei Erwachsenen [Speech Audiometric Measurements with the Simplifies Oldenburg Sentence Test OLKISA with Adults]." *Zeitschrift Für Audiologie, 55*(1), 6–13.

Buss, E., Hodge, S., Calandruccio, L., Leibold, L., & Grose, J. (2019). Masked sentence recognition in children, young adults, and older adults. *Ear And Hearing*, *40*(5), 1117-1126. https://doi.org/10.1097/aud.0000000000000692

Calamia, M., Markon, K., & Tranel, D. (2012). Scoring higher the second time around: meta-analyses of practice effects in neuropsychological assessment. *The Clinical Neuropsychologist*, *26*(4), 543-570. https://doi.org/10.1080/13854046.2012.680913

Calandruccio, L., Gomez, B., Buss, E., & Leibold, L. (2014). Development and preliminary evaluation of a pediatric Spanish–English speech perception task. *American Journal Of Audiology*, *23*(2), 158-172. https://doi.org/10.1044/2014_aja-13-0055

Cansino, S., Hernández-Ramos, E., Estrada-Manilla, C., Torres-Trejo, F., Martínez-Galindo, J., & Ayala-Hernández, M. et al. (2013). The decline of verbal and visuospatial working memory across the adult life span. *AGE*, *35*(6), 2283-2302. https://doi.org/10.1007/s11357-013-9531-1

Cansino, S., Torres-Trejo, F., Estrada-Manilla, C., Pérez-Loyda, M., Ramírez-Barajas, L., & Hernández-Ladrón-deGuevara, M. et al. (2020). Predictors of working memory maintenance and decline in older adults. *Archives Of Gerontology And Geriatrics*, *89*, 104074. https://doi.org/10.1016/j.archger.2020.104074

Cervera, T., Soler, M., Dasi, C., & Ruiz, J. (2009). Speech recognition and working memory capacity in young-elderly listeners: Effects of hearing sensitivity. *Canadian Journal Of*

Experimental Psychology/Revue Canadienne De Psychologie Expérimentale, *63*(3), 216-226. https://doi.org/10.1037/a0014321

Clopper, C., Pisoni, D., & Tierney, A. (2006). Effects of open-set and closed-set task demands on spoken word recognition. *Journal Of The American Academy Of Audiology*, *17*(05), 331-349. https://doi.org/10.3766/jaaa.17.5.4

Conrad, R., & Hull, A. (1964). Information, acoustic confusion and memory span. *British Journal Of Psychology*, *55*(4), 429-432. https://doi.org/10.1111/j.2044-8295.1964.tb00928.x

Conway, A., Kane, M., Bunting, M., Hambrick, D., Wilhelm, O., & Engle, R. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769-786. https://doi.org/10.3758/bf03196772

Conway, C., Pisoni, D., & Kronenberger, W. (2009). The importance of sound for cognitive sequencing abilities. *Current Directions In Psychological Science*, *18*(5), 275-279. https://doi.org/10.1111/j.1467-8721.2009.01651.x

Corbin, N., Bonino, A., Buss, E., & Leibold, L. (2016). Development of open-set word recognition in children. *Ear And Hearing*, *37*(1), 55-63. https://doi.org/10.1097/aud.0000000000000201

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral And Brain Sciences*, *24*(1), 87-114. https://doi.org/10.1017/s0140525x01003922

Cowan, N. (2008). Chapter 20 What are the differences between long-term, short-term, and working memory? *Progress In Brain Research*, 323-338. https://doi.org/10.1016/s0079-6123(07)00020-9

Craik, F. (1986). A functional account of age differences in memory. In F. Klix & H. Hagendorf, *Human Memory and Cognitive Capabilities, Mechanisms and Performances* (p. 409). Elsevier.

Craik, F., & Byrd, M. (1982). Aging and cognitive deficits: the role of attentional resources. In F. Craik & S. Trehub, *Aging and Cognitive Processes* (p. 191). Plenum.

Crannell, C., & Parrish, J. (1957). A comparison of immediate memory span for digits, letters, and words. *The Journal Of Psychology*, *44*(2), 319-327. https://doi.org/10.1080/00223980.1957.9713089

Cupples, L., Ching, T., Button, L., Seeto, M., Zhang, V., & Whitfield, J. et al. (2017). Spoken language and everyday functioning in 5-year-old children using hearing aids or cochlear implants. *International Journal Of Audiology*, *57*(sup2), S55-S69. https://doi.org/10.1080/14992027.2017.1370140

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal Of Verbal Learning And Verbal Behavior*, *19*(4), 450-466. https://doi.org/10.1016/s0022-5371(80)90312-6

Davidson, L., Geers, A., Hale, S., Sommers, M., Brenner, C., & Spehar, B. (2019). Effects of early auditory deprivation on working memory and reasoning abilities in verbal and visuospatial domains for pediatric cochlear implant recipients. *Ear And Hearing*, *40*(3), 517-528. https://doi.org/10.1097/aud.0000000000000629

Dawson, P., Busby, P., McKay, C., & Clark, G. (2002). Short-term auditory memory in children using cochlear implants and its relevance to receptive language. *Journal Of Speech, Language, And Hearing Research*, *45*(4), 789-801. https://doi.org/10.1044/1092-4388(2002/064)

Diefendorf, A., & Wynne, M. (2004). Paediatric Audiology: The test battery approach. In R. Kent, *The MIT encyclopaedia of communication disorders* (1st ed., pp. 520-521). MIT Press.

Dietz, A., Buschermöhle, M., Aarnisalo, A., Vanhanen, A., Hyyrynen, T., & Aaltonen, O. et al. (2014). The development and evaluation of the Finnish Matrix Sentence Test for speech intelligibility assessment. *Acta Oto-Laryngologica*, *134*(7), 728-737. https://doi.org/10.3109/00016489.2014.898185

Dietz, A., Buschermöhle, M., Sivonen, V., Willberg, T., Aarnisalo, A., Lenarz, T., & Kollmeier, B. (2015). Characteristics and international comparability of the Finnish matrix sentence test in cochlear implant recipients. *International Journal Of Audiology*, *54*(sup2), 80-87. https://doi.org/10.3109/14992027.2015.1070309

Dirks, D., Morgan, D., & Dubno, J. (1982). A procedure for quantifying the effects of noise on speech recognition. *Journal Of Speech And Hearing Disorders*, *47*(2), 114-123. https://doi.org/10.1044/jshd.4702.114

Etymōtic Research. (2005). *Bamford-Kowal-Bench Speech-in-Noise Test (Version 1.03)* [Audio CD] Elk Grove Village, IL.

Foo, C., Rudner, M., Rönnberg, J., & Lunner, T. (2007). Recognition of speech in noise with new hearing instrument compression release settings requires explicit cognitive storage and Processing Capacity. *Journal Of The American Academy Of Audiology*, *18*(07), 618-631. https://doi.org/10.3766/jaaa.18.7.8

Francart, T., van Wieringen, A., & Wouters, J. (2010). Comparison of fluctuating maskers for speech recognition tests. *International Journal Of Audiology*, *50*(1), 2-13. https://doi.org/10.3109/14992027.2010.505582

Fredriksson, S., Hammar, O., Magnusson, L., Kähäri, K., & Persson Waye, K. (2016). Validating self-reporting of hearing-related symptoms against pure-tone audiometry, otoacoustic emission, and speech audiometry. *International Journal Of Audiology*, *55*(8), 454-462. https://doi.org/10.1080/14992027.2016.1177210

Gathercole, S., Pickering, S., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2), 177-190. https://doi.org/10.1037/0012-1649.40.2.177

Glenberg, A., Bradley, M., Stevenson, J., Kraus, T., Tkachuk, M., & Gretz, A. et al. (1980). A two-process account of long-term serial position effects. *Journal Of Experimental Psychology: Human Learning And Memory*, *6*(4), 355-369. https://doi.org/10.1037/0278-7393.6.4.355

Glisky, E. (2007). Changes in Cognitive Function in Human Aging. In D. Riddle, *Brain Aging: Models, Methods, and Mechanisms.*. CRC Press/Taylor & Francis. Retrieved 13 January 2021, from https://www.ncbi.nlm.nih.gov/books/NBK3885/.

Hadjikakou, K., Petridou, L., & Stylianou, C. (2008). The academic and social inclusion of oral deaf and hard-of-hearing children in Cyprus secondary general education: investigating the perspectives of the stakeholders. *European Journal Of Special Needs Education*, *23*(1), 17-29. https://doi.org/10.1080/08856250701791211

Hagerman, B. (1976). Reliability in the determination of speech discrimination. *International Journal Of Audiology*, *5*(4), 219-228. https://doi.org/10.3109/14992027609044991

Hagerman, B. (1982). Sentences for testing speech intelligibility in noise. *Scandinavian Audiology*, *11*(2), 79-87. https://doi.org/10.3109/01050398209076203

Hagerman, B., & Hermansson, E. (2015). Speech recognition in noise in 5-year-old normal hearing children. *Canadian Journal Of Speech And Language Pathology And Audiology*, *39*(1), 52-60. https://cjslpa.ca/files/2015_CJSLPA_Vol_39/CJSLPA_Spring_2015_Vol_39_No_1.pdf#page=54.

Hall, S. J. (2006). *The Development of a New English Sentence in Noise Test and an English Number Recognition test* [Master's Thesis, University of Southampton] ePrints Soton. https://biobank.ndph.ox.ac.uk/crystal/crystal/docs/audiology_msc2006.pdf.

Hall, J., Grose, J., Buss, E., & Dev, M. (2002). Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children. *Ear And Hearing*, *23*(2), 159-165. https://doi.org/10.1097/00003446-200204000-00008

Hancock, A., LaPointe, L., Stierwalt, J., Bourgeois, M., & Zwaan, R. (2007). Computerized measures of verbal working memory performance in healthy elderly participants. *Contemporary Issues In Communication Science And Disorders*, *34*(Fall), 73-85. https://doi.org/10.1044/cicsd_34_f_73

Hasher, L., & Zacks, R. (1988). Working memory, comprehension, and aging: a review and a new view. In G. Bower, *The Psychology of Learning and Motivation* (22nd ed., p. 193). Academic Press.

Hasher, L., Zacks, R., & May, C. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher & A. Koriat, *Attention and Performance XVII* (p. 653). MIT Press.

Hawthorn, D. (2000). Possible implications of aging for interface designers. *Interacting With Computers*, *12*(5), 507-528. https://doi.org/10.1016/s0953-5438(99)00021-1

Hällgren, M., Larsby, B., Lyxell, B., & Arlinger, S. (2001). Cognitive effects in dichotic speech testing in elderly persons. *Ear And Hearing*, *22*(2), 120-129. https://doi.org/10.1097/00003446-200104000-00005

Healy, E., & Yoho, S. (2016). Difficulty understanding speech in noise by the hearing impaired: Underlying causes and technological solutions. *2016 38Th Annual International Conference Of The IEEE Engineering In Medicine And Biology Society (EMBC)*. https://doi.org/10.1109/embc.2016.7590647

Heinrich, A., Henshaw, H., & Ferguson, M. (2015). The relationship of speech intelligibility with hearing sensitivity, cognition, and perceived hearing difficulties varies for different speech perception tests. *Frontiers In Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00782

Hochmuth, S., Brand, T., Zokoll, M. A., Castro, F. Z., Wardenga, N., & Kollmeier, B. (2012). A Spanish matrix sentence test for assessing speech reception thresholds in noise. *International Journal of Audiology, 51,* 536-544. http://dx.doi.org/10.3109 /14992027.2012.670731

Holmes, J., Powell-Griner, E., Lethbridge-Cejku, M., & Heyman, K. (2009). Aging differently: Physical limitations among adults aged 50 years and over: United States, 2001-2007. *NCHS data brief*, (20), 1–8.

Hope, R. V. (2010). *Towards the Development of the New Zealand Hearing in Noise Test (NZHINT)* [Master's thesis, University of Canterbury] UC Research Repository. http://dx.doi.org/10.26021/7544

Hopkins, K., & Moore, B. (2009). The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise. *The Journal Of The Acoustical Society Of America*, *125*(1), 442-446. https://doi.org/10.1121/1.3037233

Hoppe, U., Hast, A., & Hocke, T. (2015). Audiometry-based screening procedure for cochlear implant candidacy. *Otology & Neurotology*, *36*(6), 1001-1005. https://doi.org/10.1097/mao.0000000000000730

Houben, R., Koopman, J., Luts, H., Wagener, K., van Wieringen, A., Verschuure, H., & Dreschler, W. (2014). Development of a Dutch matrix sentence test to assess speech intelligibility in noise. *International Journal Of Audiology*, *53*(10), 760-763. https://doi.org/10.3109/14992027.2014.920111

Hulme, C., Roodenrys, S., Brown, G., & Mercer, R. (1995). The role of long-term memory mechanisms in memory span. *British Journal Of Psychology*, *86*(4), 527-536. https://doi.org/10.1111/j.2044-8295.1995.tb02570.x

Humes, L. (2015). Age-related hearing loss, noise-induced hearing loss, and speech-understanding performance. *The Journal Of The Acoustical Society Of America*, *137*(4), 2248-2248. https://doi.org/10.1121/1.4920200

Humes, L., Dubno, J., Gordon-Salant, S., Lister, J., Cacace, A., & Cruickshanks, K. et al. (2012). Central presbycusis: A review and evaluation of the evidence. *Journal Of The American Academy Of Audiology*, *23*(08), 635-666. https://doi.org/10.3766/jaaa.23.8.5

Hughson, W., & Westlake, H. (1944). Manual for program outline for rehabilitation of aural casualties both military and civilian. *Transactions - American Acadamy of Opthalmology and Otolaryngology.* (suppl 48), 1-15.

Hutcherson, R., Dirks, D., & Morgan, D. (1979). Evaluation of the speech perception in noise (spin) test. *Otolaryngology–Head And Neck Surgery*, *87*(2), 239-245. https://doi.org/10.1177/019459987908700215

Hülür, G., Keye-Ehing, D., Oberauer, K., & Wilhelm, O. (2019). The effect of stimulus-response compatibility on the association of fluid intelligence and working memory with choice reaction times. *Journal Of Cognition*, *2*(1). https://doi.org/10.5334/joc.66

Iliadou, V., Ptok, M., Grech, H., Pedersen, E., Brechmann, A., & Deggouj, N. et al. (2017). A European perspective on auditory processing disorder-current knowledge and future research focus. *Frontiers In Neurology*, *8*. https://doi.org/10.3389/fneur.2017.00622

Jalilevand, N., Ebrahimipour, M., Motasaddi Zarandi, M., Kamali, M., & Fayazi, L. (2018). Comparison of verbal and non-verbal communication between deaf children with no cochlear implantation and deaf children 12-24 Months After Cochlear Implantation. *Function And Disability Journal*, *1*(3), 1-7. https://doi.org/10.30699/fdisj.1.3.1

Jansen, S., Luts, H., Wagener, K., Kollmeier, B., Del Rio, M., & Dauman, R. et al. (2012). Comparison of three types of French speech-in-noise tests: A multi-center study. *International Journal Of Audiology*, *51*(3), 164-173. https://doi.org/10.3109/14992027.2011.633568

Jenkins-Foreman (2018). *Development and Preliminary Evaluation of the University of Canterbury Paediatric Auditory-Visual Matrix Sentence Test.* [Master's thesis, University of Canterbury]. UC Research Repository. http://dx.doi.org/10.26021/8026

Jerger, J., & Hayes, D. (1977). Diagnostic speech audiometry. *Archives Of Otolaryngology - Head And Neck Surgery*, *103*(4), 216-222. https://doi.org/10.1001/archotol.1977.00780210072008

Jerger, J., Jerger, S., & Pirozzolo, F. (1991). Correlational analysis of speech audiometric scores, hearing loss, age, and cognitive abilities in the elderly. *Ear And Hearing*, *12*(2), 103-109. https://doi.org/10.1097/00003446-199104000-00004

Jerger, S., & Jerger, J. (1982). Pediatric speech intelligibility test. *Ear And Hearing*, *3*(6), 325-334. https://doi.org/10.1097/00003446-198211000-00007

Jonides, J., Lewis, R., Nee, D., Lustig, C., Berman, M., & Moore, K. (2008). The mind and brain of short-term memory. *Annual Review Of Psychology*, *59*(1), 193-224. https://doi.org/10.1146/annurev.psych.59.103006.093615

Kahana, M., & Loftus, G. (1999). Response time versus accuracy in human memory. In R. Sternberg, *The Nature of Cognition* (pp. 323-375). The MIT Press. Retrieved 18 January 2021, from https://memory.psych.upenn.edu/files/pubs/KahaLoft99.pdf.

Kalikow, D., Stevens, K., & Elliott, L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal Of The Acoustical Society Of America*, *61*(5), 1337-1351. https://doi.org/10.1121/1.381436

Kelly-Campbell, R., & Lessoway, K. (2015). Hearing aid and hearing assistance technology use in Aotearoa/New Zealand. *International Journal Of Audiology*, *54*(5), 308-315. https://doi.org/10.3109/14992027.2014.979952

Kelly-Campbell, R., Thomas, K., & McMillan, A. (2014). Audiological outcomes for adults with mild hearing impairment: A pilot study. *Speech, Language And Hearing*, *18*(2), 74-82. https://doi.org/10.1179/2050572814y.0000000052

Killion, M. (2002). New thinking on hearing in noise: A generalized articulation index. *Seminars In Hearing*, *23*(1), 057-076. https://doi.org/10.1055/s-2002-24976

Killion, M., Niquette, P., Gudmundsen, G., Revit, L., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *The Journal Of The Acoustical Society Of America*, *116*(4), 2395-2405. https://doi.org/10.1121/1.1784440

King, C., Warrier, C., Hayes, E., & Kraus, N. (2002). Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. *Neuroscience Letters*, *319*(2), 111-115. https://doi.org/10.1016/s0304-3940(01)02556-3

King, S. M. (2010). *Development and Evaluation of a New Zealand Digit Triplet Test for Auditory Screening* [Master's Thesis, University of Canterbury]. UC Research Repository. http://dx.doi.org/10.26021/7820

Kirk, K., Diefendorf, A., Pisoni, D., & Robbins, A. (1997). Assessing speech perception in children. In L. Mendel & J. Danhauer, *Audiologic Evaluation and Management and Speech Perception Assessment* (pp. 101-132). Singular Publishing Company, Inc.

Klencklen, G., Banta Lavenex, P., Brandner, C., & Lavenex, P. (2017). Working memory decline in normal aging: Is it really worse in space than in color?. *Learning And Motivation*, *57*, 48-60. https://doi.org/10.1016/j.lmot.2017.01.007

Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M., Uslar, V., Brand, T., & Wagener, K. (2015). The multilingual matrix test: Principles, applications, and comparison across languages: A review. *International Journal Of Audiology*, *54*(sup2), 3-16. https://doi.org/10.3109/14992027.2015.1020971

Konkle, D., & Rintelmann, W. (1983). Introduction to speech audiometry. In D. Konkle & W. Rintelmann, *Principles of Speech Audiometry* (pp. 1-10). University Park Press.

Kosky, C., & Boothroyd, A. (2003). Validation of an on-line implementation of the imitative test of speech pattern contrast perception (IMSPAC). *Journal Of The American Academy Of Audiology*, *14*(02), 072-083. https://doi.org/10.3766/jaaa.14.2.3

Lang-Roth, R. (2014). Hearing impairment and language delay in infants: Diagnostics and genetics. *GMS Current Topics In Otorhinolaryngology - Head And Neck Surgery*, *13*, Doc05. https://doi.org/10.3205/cto000108

Legris, E., Gomot, M., Charpentier, J., Aoustin, J., Aussedat, C., & Bakhos, D. (2018). Assessment of auditory discrimination in hearing-impaired patients. *European Annals Of Otorhinolaryngology, Head And Neck Diseases*, *135*(5), 335-339. https://doi.org/10.1016/j.anorl.2018.04.004

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal Of The Acoustical Society Of America*, *49*(2B), 467-477. https://doi.org/10.1121/1.1912375

Lewis, D., Schmid, K., O'Leary, S., Spalding, J., Heinrichs-Graham, E., & High, R. (2016). Effects of noise on speech recognition and listening effort in children with normal hearing and children with mild bilateral or unilateral hearing loss. *Journal Of Speech, Language, And Hearing Research*, *59*(5), 1218-1232. https://doi.org/10.1044/2016_jslhr-h-15-0207

Litovsky, R. (2005). Speech intelligibility and spatial release from masking in young children. *The Journal Of The Acoustical Society Of America*, *117*(5), 3091-3099. https://doi.org/10.1121/1.1873913

Llorach, G., Kirschner, F., Grimm, G., Zokoll. M., Wagener, K., & Hohmann, V. (2019). Development and evaluation of video recordings for the OLSA Matrix Sentence Test. *arXiv.* https://arxiv.org/abs/1912.04700

Maclagan, M., & Hay, J. (2007). Getting fed up with our feet: Contrast maintenance and the New Zealand English "short" front vowel shift. *Language Variation And Change*, *19*(01). https://doi.org/10.1017/s0954394507070020

Maclennan-Smith, F., Swanepoel, D., & Hall, J. (2012). Validity of diagnostic pure-tone audiometry without a sound-treated environment in older adults. *International Journal Of Audiology*, *52*(2), 66-73. https://doi.org/10.3109/14992027.2012.736692

Madell, J., & Flexer, C. (2014). *Pediatric Audiology: Diagnosis, Technology, and Management* (2nd ed., pp. 107-108). Thieme.

Marshall, P., & Werder, P. (1972). The effects of the elimination of rehearsal on primacy and recency. *Journal Of Verbal Learning And Verbal Behavior*, *11*(5), 649-653. https://doi.org/10.1016/s0022-5371(72)80049-5

May, C., Hasher, L., & Kane, M. (1999). The role of interference in memory span. *Memory & Cognition*, *27*(5), 759-767. https://doi.org/10.3758/bf03198529

McArdle, R., & Hnath-Chisolm, T. (2015). Speech Audiometry. In J. Katz, M. Chasin, K. English, L. Hood & K. Tillery, *Handbook of clinical audiology* (7th ed., pp. 61-72). Wolters Kluwer Health.

McArdle, R., Wilson, R., & Burks, C. (2005). Speech recognition in multitalker babble using digits, words, and sentences. *Journal Of The American Academy Of Audiology*, *16*(09), 726-739. https://doi.org/10.3766/jaaa.16.9.9

McGaffin, A. J. (2007). *Development of a monosyllabic adaptive speech test for the identification of central auditory processing disorder* [Master's Thesis, University of Canterbury]. UC Research Repository. http://dx.doi.org/10.26021/7892

McGarrigle, R., Munro, K., Dawes, P., Stewart, A., Moore, D., Barry, J., & Amitay, S. (2014). Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *International Journal Of Audiology*, *53*(7), 433-445. https://doi.org/10.3109/14992027.2014.890296

McNab, F., Zeidman, P., Rutledge, R., Smittenaar, P., Brown, H., Adams, R., & Dolan, R. (2015). Age-related changes in working memory and the ability to ignore distraction. *Proceedings Of The National Academy Of Sciences*, *112*(20), 6515-6518. https://doi.org/10.1073/pnas.1504162112

Meister, H. (2016). Speech audiometry, speech perception, and cognitive functions. *HNO*, *65*(S1), 1-4. https://doi.org/10.1007/s00106-016-0250-7

Meister, H., Rählmann, S., Lemke, U., & Besser, J. (2018). Verbal response times as a potential indicator of cognitive load during conventional speech audiometry with matrix sentences. *Trends In Hearing*, *22*, 233121651879325. https://doi.org/10.1177/2331216518793255

Mendel, L. (2008). Current considerations in pediatric speech audiometry. *International Journal Of Audiology*, *47*(9), 546-553. https://doi.org/10.1080/14992020802252261

Miller, G. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81-97. https://doi.org/10.1037/h0043158

Mills, C., Potenza, M., Fremer, J., & Ward, W. (2014). *Computer-based testing* (1st ed., p. 249). Psychology Press.

Ministry of Health. (2016). Universal newborn hearing screening and early intervention programme: National policy and quality standards: Diagnostic and amplification protocols. Wellington: Ministry of Health. https://www.nsu.govt.nz/system/files/resources/unhseip-policy-quality-standards-diagnostic-amplification-protocols-jan16.pdf

Murata, A., & Iwase, H. (2005). Usability of touch-panel interfaces for older adults. *Human Factors: The Journal Of The Human Factors And Ergonomics Society*, *47*(4), 767-776. https://doi.org/10.1518/001872005775570952

Murdock, B. (1962). The serial position effect of free recall. *Journal Of Experimental Psychology*, *64*(5), 482-488. https://doi.org/10.1037/h0045106

Murdock, B., & Okada, R. (1970). Interresponse times in single-trial free recall. *Journal Of Experimental Psychology*, *86*(2), 263-267. https://doi.org/10.1037/h0029993

Neumann, K., Baumeister, N., Baumann, U., Sick, U., Euler, H., & Weißgerber, T. (2012). Speech audiometry in quiet with the Oldenburg Sentence Test for Children. *International Journal Of Audiology*, *51*(3), 157-163. https://doi.org/10.3109/14992027.2011.633935

Newman, C., & Sandridge, S. (2004). Hearing loss is often undiscovered, but screening is easy. *Cleveland Clinic Journal Of Medicine*, *71*(3), 225-232. https://doi.org/10.3949/ccjm.71.3.225

New Zealand Audiological Society (NZAS). (2015). Professional practice standards *Part B Clinical Practice*. Auckland, New Zealand: New Zealand Audiological Society.

New Zealand Audiological Society (NZAS). (2016). *NZAS Best Practice Guidelines: Speech Audiometry in Diagnostic Hearing Assessment for Young Persons and Adults* (pp. 3-13). Auckland: New Zealand Audiological Society.

New Zealand Audiological Society (NZAS). (2020). *Adult Pure Tone Audiometry Best Practice Guidelines* (pp. 9-10). Auckland: New Zealand Audiological Society.

Ng, S., Meston, C., Scollie, S., & Seewald, R. (2011). Adaptation of the BKB-SIN Test for use as a pediatric aided outcome measure. *Journal Of The American Academy Of Audiology*, *22*(06), 375-386. https://doi.org/10.3766/jaaa.22.6.6

Nilsson, M., Soli, S., & Sullivan, J. (1994). Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal Of The Acoustical Society Of America*, *95*(2), 1085-1099. https://doi.org/10.1121/1.408469

Norrix, L. (2015). Hearing thresholds, minimum response levels, and cross-check measures in pediatric audiology. *American Journal Of Audiology*, *24*(2), 137-144. https://doi.org/10.1044/2015_aja-14-0095

Olusanya, B., Neumann, K., & Saunders, J. (2014). The global burden of disabling hearing impairment: a call to action. *Bulletin Of The World Health Organization*, *92*(5), 367-373. https://doi.org/10.2471/blt.13.128728

Ostergard, C. (1983). Factors influencing the validity and reliability of speech audiometry. *Seminars In Hearing*, *4*(03), 221-239. https://doi.org/10.1055/s-0028-1091427

Ozimek, E., Kutzner, D., & Libiszewski, P. (2012). Speech intelligibility tested by the Pediatric Matrix Sentence test in 3–6year old children. *Speech Communication*, *54*(10), 1121-1131. https://doi.org/10.1016/j.specom.2012.06.001

Ozimek, E., Kutzner, D., Sęk, A., & Wicher, A. (2009). Polish sentence tests for measuring the intelligibility of speech in interfering noise. *International Journal Of Audiology*, *48*(7), 433-443. https://doi.org/10.1080/14992020902725521

Ozimek, E., Warzybok, A., & Kutzner, D. (2010). Polish sentence matrix test for speech intelligibility measurement in noise. *International Journal Of Audiology*, *49*(6), 444-454. https://doi.org/10.3109/14992021003681030

O'Beirne, G., McGaffin, A., & Rickard, N. (2012). Development of an adaptive low-pass filtered speech test for the identification of auditory processing disorders. *International Journal Of Pediatric Otorhinolaryngology*, *76*(6), 777-782. https://doi.org/10.1016/j.ijporl.2012.02.039

O'Beirne, G. A., Trouson, R. H., McClelland, A. D., Jamaluddin, S., Maclagan, M. A. (2015). Development of an auditory-visual matrix sentence test in New Zealand English. *12th European Federation of Audiological Societies Congress.* Istanbul, Turkey, 30 May 2015.

Park, D. (2000). The basic mechanisms accounting for age-related decline in cognitive function. In D. Park & N. Schwarz, *Cognitive Aging: A Primer* (p. 3). Psychology Press.

Park, D., & Hedden, T. (2001). Working memory and aging. In M. Naveh-Benjamin, M. Moscovitch & H. Roediger, *Perspectives on Human Memory and Cognitive Aging: Essays in Honor of Fergus Craik* (p. 148). Psychology Press.

Patel, H. C., Moitra, M., Modi, A., Contractor, J., & Kantharia, S. L. (2014). Impact of hearing loss on daily life style and schooling among children between 5 and 15 years ace-group. *National Journal of Community Medicine, 5*(1), 73-76.

Percy-Smith, L., Hallstrøm, M., Josvassen, J., Mikkelsen, J., Nissen, L., Dieleman, E., & Cayé-Thomasen, P. (2018). Differences and similarities in early vocabulary development between children with hearing aids and children with cochlear implant enrolled in 3-year auditory verbal intervention. *International Journal Of Pediatric Otorhinolaryngology*, *108*, 67-72. https://doi.org/10.1016/j.ijporl.2018.02.030

Peters, R., & Moore, B. (1992). Auditory filter shapes at low center frequencies in young and elderly hearing-impaired subjects. *The Journal Of The Acoustical Society Of America*, *91*(1), 256-266. https://doi.org/10.1121/1.402769

Peters, R., Moore, B., & Baer, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *The Journal Of The Acoustical Society Of America*, *103*(1), 577-587. https://doi.org/10.1121/1.421128

Pichora-Fuller, M., Schneider, B., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal Of The Acoustical Society Of America*, *97*(1), 593-608. https://doi.org/10.1121/1.412282

Pillai, R., & Yathiraj, A. (2017). Auditory, visual and auditory-visual memory and sequencing performance in typically developing children. *International Journal Of Pediatric Otorhinolaryngology*, *100*, 23-34. https://doi.org/10.1016/j.ijporl.2017.06.010

Pisoni, D., & Cleary, M. (2003). Measures of working memory span and verbal rehearsal speed in deaf children after cochlear implantation. *Ear And Hearing*, *24*(Supplement), 106S-120S. https://doi.org/10.1097/01.aud.0000051692.05140.8e

Pisoni, D., Conway, C., Kronenberger, W., Horn, D., Karpicke, J., & Henning, S. (2008). Efficacy and effectiveness of cochlear implants in deaf children. In M. Marschark & P.

Hauser, *Deaf cognition: Foundations and Outcomes* (pp. 52-101). Oxford University Press.

Pliatsikas, C., Veríssimo, J., Babcock, L., Pullman, M., Glei, D., & Weinstein, M. et al. (2018). Working memory in older adults declines with age, but is modulated by sex and education. *Quarterly Journal Of Experimental Psychology*, *72*(6), 1308-1327. https://doi.org/10.1177/1747021818791994

Plomp, R., & Mimpen, A. (1979). Improving the reliability of testing the speech reception threshold for sentences. *International Journal Of Audiology*, *18*(1), 43-52. https://doi.org/10.3109/00206097909072618

Poirier, M., & Saint-Aubin, J. (1996). Immediate serial recall, word frequency, item identity and item position. *Canadian Journal Of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale*, *50*(4), 408-412. https://doi.org/10.1037/1196-1961.50.4.408

Portnuff, C., & Bell, B. (2019). Effective use of Speech-in-Noise Testing in the Clinic. *The Hearing Journal*, *72*(5), 40. https://doi.org/10.1097/01.hj.0000559502.51932.b1

Puglisi, G., di Berardino, F., Montuschi, C., Sellami, F., Albera, A., & Zanetti, D. et al. (2021). Evaluation of Italian Simplified Matrix Test for speech-recognition measurements in noise. *Audiology Research*, *11*(1), 73-88. https://doi.org/10.3390/audiolres11010009

Puglisi, G., Warzybok, A., Hochmuth, S., Visentin, C., Astolfi, A., Prodi, N., & Kollmeier, B. (2015). An Italian matrix sentence test for the evaluation of speech intelligibility in noise. *International Journal Of Audiology*, *54*(sup2), 44-50. https://doi.org/10.3109/14992027.2015.1061709

Purdy, S., Kelly, A., & Davies, M. (2002). Auditory brainstem response, middle latency response, and late cortical evoked potentials in children with learning disabilities. *Journal*

Of The American Academy Of Audiology, *13*(07), 367-382. https://doi.org/10.1055/s-0040-1715981

Reuter-Lorenz, P., & Sylvester, C. (2005). The cognitive neuroscience of working memory and aging. In R. Cabeza, L. Nyberg & D. Park, *Cognitive Neuroscience of Aging* (p. 186). Oxford University Press.

Roland, L., Fischer, C., Tran, K., Rachakonda, T., Kallogjeri, D., & Lieu, J. (2016). Quality of life in children with hearing impairment. *Otolaryngology–Head And Neck Surgery*, *155*(2), 208-219. https://doi.org/10.1177/0194599816640485

Ross, M., & Lerman, J. (1970). A Picture dentification Test for hearing-impaired children. *Journal Of Speech And Hearing Research*, *13*(1), 44-53. https://doi.org/10.1044/jshr.1301.44

Rowe, A., Titterington, J., Holmes, J., Henry, L., & Taggart, L. (2019). Interventions targeting working memory in 4–11 year olds within their everyday contexts: A systematic review. *Developmental Review*, *52*, 1-23. https://doi.org/10.1016/j.dr.2019.02.001

Rudner, M., Foo, C., Rönnberg, J., & Lunner, T. (2009). Cognition and aided speech recognition in noise: Specific role for cognitive factors following nine-week experience with adjusted compression settings in hearing aids. *Scandinavian Journal Of Psychology*, *50*(5), 405-418. https://doi.org/10.1111/j.1467-9450.2009.00745.x

Rudner, M., Lunner, T., Behrens, T., Thorén, E., & Rönnberg, J. (2012). Working memory capacity may influence perceived effort during aided speech recognition in noise. *Journal Of The American Academy Of Audiology*, *23*(08), 577-589. https://doi.org/10.3766/jaaa.23.7.7

Rudner, M., Rönnberg, J., & Lunner, T. (2011). Working memory supports listening in noise for persons with hearing impairment. *Journal Of The American Academy Of Audiology*, *22*(03), 156-167. https://doi.org/10.3766/jaaa.22.3.4

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal Of Experimental Psychology*, *89*(1), 63-77. https://doi.org/10.1037/h0031185

Rundus, D. (1980). Maintenance rehearsal and long-term recency. *Memory & Cognition*, *8*(3), 226-230. https://doi.org/10.3758/bf03197610

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., & Lyxell, B. et al. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers In Systems Neuroscience*, *7*. https://doi.org/10.3389/fnsys.2013.00031

Salthouse, T. (1995). Processing capacity and its role on the relations between age and memory. In F. Weinert & W. Schneider, *Memory Performance and Competencies: Issues in Growth and Development* (p. 111). Erlbaum.

Sarant, J., Holt, C., Dowell, R., Rickards, F., & Blamey, P. (2008). Spoken language development in oral preschool children with permanent childhood deafness. *Journal Of Deaf Studies And Deaf Education*, *14*(2), 205-217. https://doi.org/10.1093/deafed/enn034

Schafer, C. (2010). Speech perception in noise measures for children: a critical review and case studies. *Journal of Educational Audiology, 16,* 4-15.

Schlauch, R., & Nelson, P. (2015). Puretone evaluation. In Katz, J., Chasin, M., English, K., Hood, L., & Tillery, K. *Handbook of Clinical Audiology* (7th ed., pp. 40-42). Philadelphia: Wolters Kluwer Health.

Schnipke, D. L. (1995). *Assessing speededness in computer-based tests using item response times (Unpublished doctoral dissertation). Johns Hopkins University.*

Schnipke, D. L., & Scrams, D. J. (2002). Exploring issues of examinee behavior: Insights gained from response-time analyses. In C. N. Mills, M. T. Potenza, J. J. Fremer, & W. C. Ward (Eds.), *Computer-based testing: Building the foundation for future assessments* (pp. 237–266). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Service, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *The Quarterly Journal Of Experimental Psychology Section A*, *51*(2), 283-304. https://doi.org/10.1080/713755759

Smith, M., Sharit, J., & Czaja, S. (1999). Aging, motor control, and the performance of computer mouse tasks. *Human Factors: The Journal Of The Human Factors And Ergonomics Society*, *41*(3), 389-396. https://doi.org/10.1518/001872099779611102

Soli, S., & Wong, L. (2008). Assessment of speech intelligibility in noise with the Hearing in Noise Test. *International Journal Of Audiology*, *47*(6), 356-361. https://doi.org/10.1080/14992020801895136

Sommers, M., Tye-Murray, N., & Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear And Hearing*, *26*(3), 263-275. https://doi.org/10.1097/00003446-200506000-00003

Sperry, J.L., Wiley, T.L., & Chial, M.R. (1997). Word recognition performance in various background competitors. *Journal of the American Academy of Audiology, 8*(2), 71-80. https://www.audiology.org/sites/default/files/journal/JAAA_08_02_01.pdf

Statistics New Zealand. (2014). *Disability Survey: 2013* (p. 6). Tauranga: Statistics New Zealand. From https://www.stats.govt.nz/information-releases/disability-survey-2013

Stelmachowicz, P., Pittman, A., Hoover, B., Lewis, D., & Moeller, M. (2004). The importance of high-frequency audibility in the speech and language development of children with hearing loss. *Archives Of Otolaryngology–Head & Neck Surgery*, *130*(5), 556. https://doi.org/10.1001/archotol.130.5.556

Stevens, G., Flaxman, S., Brunskill, E., Mascarenhas, M., Mathers, C., & Finucane, M. (2011). Global and regional hearing impairment prevalence: an analysis of 42 studies in 29 countries. *European Journal Of Public Health*, *23*(1), 146-152. https://doi.org/10.1093/eurpub/ckr176

Stevenson, J., Kreppner, J., Pimperton, H., Worsfold, S., & Kennedy, C. (2015). Emotional and behavioural difficulties in children and adolescents with hearing impairment: a systematic review and meta-analysis. *European Child & Adolescent Psychiatry*, *24*(5), 477-496. https://doi.org/10.1007/s00787-015-0697-1

Stuart, A. (2005). Development of auditory temporal resolution in cchool-age children revealed by word recognition in continuous and interrupted noise. *Ear And Hearing*, *26*(1), 78-88. https://doi.org/10.1097/00003446-200502000-00007

Tao, D., Fu, Q., Galvin, J., & Yu, Y. (2017). The development and validation of the Closed-set Mandarin Sentence (CMS) test. *Speech Communication*, *92*, 125-131. https://doi.org/10.1016/j.specom.2017.06.008

Theunissen, M., Swanepoel, D., & Hanekom, J. (2009). Sentence recognition in noise: Variables in compilation and interpretation of tests. *International Journal Of Audiology*, *48*(11), 743-757. https://doi.org/10.3109/14992020903082088

Tremblay, K., Kraus, N., Carrell, T. D., & McGee, T. (1997). Central auditory system plasticity: Generalization to novel stimuli following listening training. Special Education and

Communication Disorders Faculty Publication. Paper 31. http://digitalcommons.unl.edu/specedfacpub/31

Trounson, R. H. (2012). *Development of the UC Auditory-Visual Matrix Sentence Test* [Master's Thesis, University of Canterbury]. UC Research Repository. http://dx.doi.org/10.26021/6028

Tye-Murray, N., Sommers, M., & Spehar, B. (2007). Auditory and visual lexical neighborhoods in audiovisual speech perception. *Trends In Amplification*, *11*(4), 233-241. https://doi.org/10.1177/1084713807307409

Uhler, K., Biever, A., & Gifford, R. (2016). Method of speech stimulus presentation impacts pediatric speech recognition. *Otology & Neurotology*, *37*(2), e70-e74. https://doi.org/10.1097/mao.0000000000000911

van Rooij, J., & Plomp, R. (1990). Auditive and cognitive factors in speech perception by elderly listeners. II: Multivariate analyses. *The Journal Of The Acoustical Society Of America*, *88*(6), 2611-2624. https://doi.org/10.1121/1.399981

Wagener, K., & Brand, T. (2005). Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters. *International Journal Of Audiology*, *44*(3), 144-156. https://doi.org/10.1080/14992020500057517

Wagener, K. C., B. Kollmeier. 2005. "Evaluation des Oldenburger Satztests mit Kindern und Oldenburger Kinder-Satztest (Evaluation of the Oldenburg Sentence Test with Children and the Oldenburg Children's Sentence Test)." Zeitschrift Fur Audiologie € 44: 134–143.

Wagener, K, C., Kühnel, V., & Kollmeier, B. (1999a). Entwicklung und Evaluation eines Satztests für die Deutsche Sprache I: Design des Oldenburger Satztests Development and evaluation of a sentence test for the German language I: Design of the Oldenburg sentence test. *Zeitschrift für Audiologie, 38(2),* 4-15.

Wagener, K. C., Brand, T., & Kollmeier, B. (1999b). Entwicklung und Evaluation eines Satztests für die Deutsche Sprache II: Optimierung des Oldenburger Satzetests (Development and evaluation of a sentence test for the German language II: Optimization of the Oldenburg sentence test). *Zeitschrift für Audiologie, 38(2),* 44-56.

Wagener, K. C., Brand, T., & Kollmeier, B. (1999c). Entwicklung und Evaluation eines Satztests für die Deutsche Sprache III: Evaluation des Oldenburger Satztests (Development and evaluation of a sentence test for the German language III: Evaluation of the Oldenburg sentence test). *Zeitschrift für Audiologie, 38(2),* 86-95.

Wagener, K., Josvassen, J., & Ardenkjær, R. (2003). Design, optimization and evaluation of a Danish sentence test in noise: Diseño, optimización y evaluación de la prueba Danesa de frases en ruido. *International Journal Of Audiology*, *42*(1), 10-17. https://doi.org/10.3109/14992020309056080

Walker, N., Millians, J., Worden, A. (1996). Mouse accelerations and performance of older computer users. In *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (pp. 151–154). Santa Monica, CA: Human Factors and Ergonomics Society. Retrieved from http://pro.sagepub.com.ezproxy.canterbury.ac.nz/content/40/3/151.short Google Scholar

Walker, E., Holte, L., McCreery, R., Spratford, M., Page, T., & Moeller, M. (2015). The influence of hearing aid use on outcomes of children with mild hearing loss. *Journal Of*

Speech, Language, And Hearing Research, *58*(5), 1611-1625. https://doi.org/10.1044/2015_jslhr-h-15-0043

Warzybok, A., Zokoll, M., Wardenga, N., Ozimek, E., Boboshko, M., & Kollmeier, B. (2015). Development of the Russian matrix sentence test. *International Journal Of Audiology*, *54*(sup2), 35-43. https://doi.org/10.3109/14992027.2015.1020969

Wilson, R., McArdle, R., & Smith, S. (2007). An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. *Journal Of Speech, Language, And Hearing Research*, *50*(4), 844-856. https://doi.org/10.1044/1092-4388(2007/059)

Wilson, R., Farmer, N., Gandhi, A., Shelburne, E., & Weaver, J. (2010). Normative data for the Words-in-Noise Test for 6- to 12-year-old children. *Journal Of Speech, Language, And Hearing Research*, *53*(5), 1111-1121. https://doi.org/10.1044/1092-4388(2010/09-0270)

Wingfield, A., Amichetti, N., & Lash, A. (2015). Cognitive aging and hearing acuity: modeling spoken language comprehension. *Frontiers In Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00684

Wise, S. (2006). An investigation of the differential effort received by items on a Low-Stakes Computer-Based Test. *Applied Measurement In Education*, *19*(2), 95-114. https://doi.org/10.1207/s15324818ame1902_2

Wise, S., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in Computer-Based Tests. *Applied Measurement In Education*, *18*(2), 163-183. https://doi.org/10.1207/s15324818ame1802_2

Wong, L., Chen, Y., & Leung, K. (2019). The Cantonese earing in Noise Test for children. *Trends In Hearing*, *23*, 233121651983712. https://doi.org/10.1177/2331216519837128

Yau, J. (2018). *Normative data for the UCAST-FW test of APD in children.* [Master's thesis, University of Canterbury]. UC Research Repository. http://dx.doi.org/10.26021/7354

Zeng, F., & Liu, S. (2006). Speech Perception in Individuals With Auditory Neuropathy. *Journal Of Speech, Language, And Hearing Research*, *49*(2), 367-380. https://doi.org/10.1044/1092-4388(2006/029)

Zokoll, M., Fidan, D., Türkyılmaz, D., Hochmuth, S., Ergenç, İ., Sennaroğlu, G., & Kollmeier, B. (2015). Development and evaluation of the Turkish matrix sentence test. *International Journal Of Audiology*, *54*(sup2), 51-61. https://doi.org/10.3109/14992027.2015.1074735

Zokoll, M., Hochmuth, S., Warzybok, A., Wagener, K., Buschermöhle, M., & Kollmeier, B. (2013). Speech-in-Noise Tests for multilingual hearing screening and diagnostics 1. *American Journal Of Audiology*, *22*(1), 175-178. https://doi.org/10.1044/1059-0889(2013/12-0061)

Appendices

APPENDIX A:

Sentence Lists

A.1 Sentence lists used for the UCAMST, showing sentences and corresponding sentence number identifiers.

List 1		List 2		List 3		List 4	
41	Kathy kept twelve green mugs	51	Amy sees twelve cheap spoons	61	Hannah has those red mugs	71	David sold nine red bikes
42	Peter has three good toys	52	Rachel sold nine new books	62	Peter gives twelve cheap toys	72	Peter bought six big ships
43	Oscar sees those red ships	53	Oscar wants three red toys	63	Thomas wants four small bikes	73	Thomas wants some good mugs
44	Thomas got eight cheap books	54	David has six good coats	64	David got ten dark coats	74	Amy likes four new books
45	David bought two big hats	55	Peter got those green shoes	65	Kathy kept six large spoons	75	Oscar got those green spoons
46	Sophie wins ten new shoes	56	Sophie likes two large hats	66	Rachel bought nine big shirts	76	William kept two dark hats
47	Amy sold six small bikes	57	William gives some dark shirts	67	Amy sold eight old shoes	77	Hannah sees twelve large shirts
48	Hannah likes some large shirts	58	Thomas kept ten small ships	68	Oscar likes some new ships	78	Rachel wins eight old coats
49	Rachel gives nine dark spoons	59	Kathy bought four big mugs	69	William wins three good books	79	Kathy has three small toys
50	William wants four old coats	60	Hannah wins eight old bikes	70	Sophie sees two green hats	80	Sophie gives ten cheap shoes
List 5		**List 6**		**List 7**		**List 8**	
81	Oscar gives six dark coats	91	Hannah gives those green hats	101	Hannah got those large shoes	111	Oscar wants twelve dark shoes
82	Hannah sees ten small ships	92	Sophie has two dark spoons	102	Thomas wants three small books	112	David kept six red ships
83	William wins two red hats	93	Thomas sees some old shirts	103	Oscar sold some dark shirts	113	Rachel got nine cheap hats
84	Sophie has nine cheap spoons	94	Peter sold six small coats	104	William sees twelve new ships	114	Thomas gives some green spoons
85	Thomas wants some large shoes	95	William likes three good shoes	105	Amy bought eight big bikes	115	Hannah wins two small bikes
86	Amy got eight good toys	96	David bought nine big ships	106	Peter gives ten cheap toys	116	Amy sees ten old coats
87	Rachel bought four big mugs	97	Amy kept twelve new bikes	107	Rachel wins four old coats	117	Peter has those large toys
88	David likes those green shirts	98	Rachel wins ten large mugs	108	Sophie has six good spoons	118	Sophie bought three big shirts
89	Peter sold three old books	99	Kathy wants four red toys	109	David likes two red mugs	119	William sold four good mugs
90	Kathy kept twelve new bikes	100	Oscar got eight cheap books	110	Kathy kept nine green hats	120	Kathy likes eight new books
List 9		**List 10**		**List 11**		**List 12**	
121	Amy gives twelve dark coats	131	Thomas likes two small spoons	141	David sold three large coats	151	Rachel gives those cheap spoons
122	David wins those cheap shirts	132	Kathy got some cheap shoes	142	Rachel has twelve red shoes	152	Kathy likes three good books
123	Kathy sold nine red books	133	Rachel wins three red mugs	143	Hannah gives six dark mugs	153	Oscar has twelve old coats
124	William has some new spoons	134	Oscar kept six green ships	144	Thomas sees eight small ships	154	Hannah sees nine new bikes
125	Thomas sees eight small hats	135	Sophie bought ten big shirts	145	Oscar likes some new shirts	155	Peter got some green mugs
126	Rachel got two good toys	136	Peter gives eight good toys	146	Sophie got nine cheap hats	156	Amy wants four red toys
127	Oscar kept six green mugs	137	Hannah sold those large bikes	147	Amy wants those green toys	157	William wins two dark shoes
128	Hannah likes three large shoes	138	William has nine old books	148	Kathy wins four old books	158	Sophie kept eight large shirts
129	Peter bought four big ships	139	Amy sees four new coats	149	Peter bought ten big spoons	159	Thomas bought ten big ships
130	Sophie wants ten old bikes	140	David wants twelve dark hats	150	William kept two good bikes	160	David sold six small hats
List 13		**List 14**		**List 15**		**List 16**	
161	William sold eight old mugs	171	Hannah gives some old spoons	181	Peter wins nine green spoons	191	Peter got three dark toys
162	Rachel got six dark coats	172	Thomas sees those green ships	182	Oscar has twelve large shoes	192	Rachel sold four red shoes
163	Kathy kept three small bikes	173	David wants twelve red mugs	183	Amy gives ten dark toys	193	Amy sees ten new bikes
164	Peter wins ten green toys	174	Rachel has eight dark books	184	David kept six good hats	194	Kathy likes some good mugs
165	Hannah gives those red shirts	175	William kept six good shirts	185	Thomas likes three new books	195	William kept those large shirts
166	David wants four good books	176	Sophie wins two small bikes	186	Sophie wants those red shirts	196	Thomas wants twelve small coats
167	Oscar sees twelve new shoes	177	Kathy sold three new toys	187	Kathy sold some small bikes	197	Sophie has two green books
168	Amy bought nine big ships	178	Peter got nine cheap shoes	188	Rachel got four cheap coats	198	Hannah gives nine cheap hats
169	Thomas has some cheap hats	179	Amy bought four big hats	189	William sees eight old ships	199	David wins eight old spoons
170	Sophie likes two large spoons	180	Oscar likes ten large coats	190	Hannah bought two big mugs	200	Oscar bought six big ships

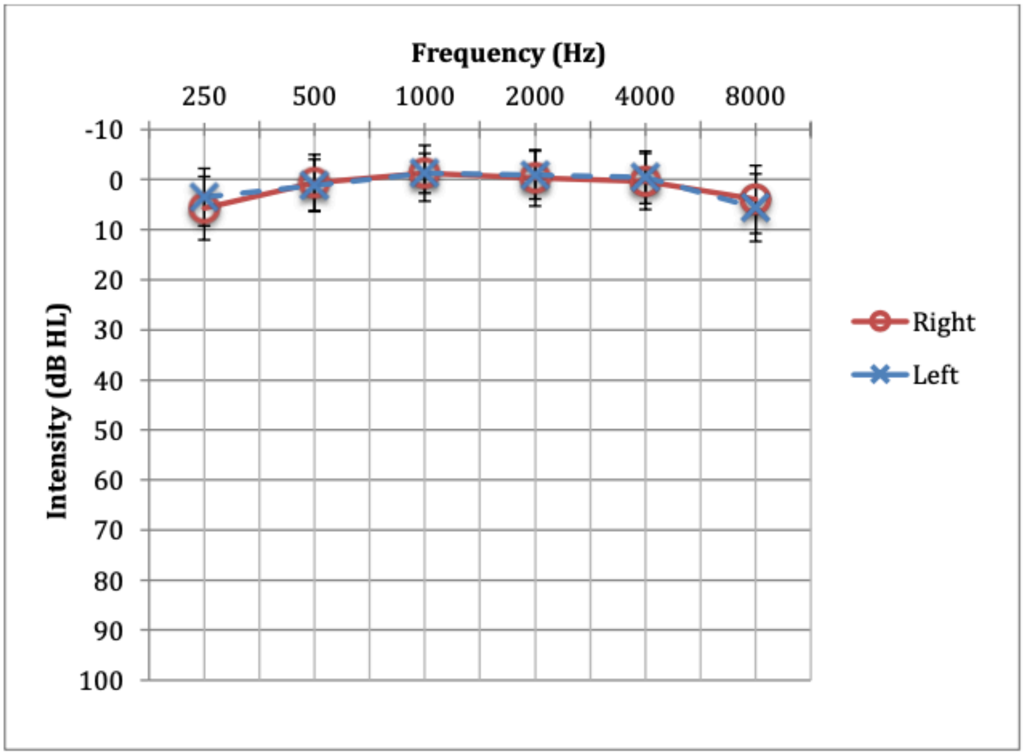
A.2 Sentence lists used for the UCAMST-P, showing sentences and corresponding sentence number identifiers.

List 1		**List 2**		**List 3**		**List 4**	
2	eight big books	21	eight old hats	13	eight new bikes	10	eight green shoes
47	nine green spoons	28	eight red shoes	69	nine small hats	23	eight old spoons
54	nine new toys	61	nine red bikes	94	ten old shoes	52	nine new shoes
88	ten new shoes	81	ten green hats	101	ten red spoons	65	nine red spoons
95	ten old spoons	132	three old toys	110	three big books	89	ten new spoons
138	three red toys	139	three small bikes	117	three green hats	93	ten old hats
156	twelve green toys	150	twelve big toys	158	twelve new books	109	three big bikes
177	twelve small hats	161	twelve new spoons	172	twelve red shoes	128	three old books
202	two old shoes	194	two new books	192	two green toys	174	twelve red toys
207	two red hats	214	two small shoes	203	two old spoons	180	twelve small toys
						213	two small hats
List 5		**List 6**		**List 7**		**List 8**	
29	eight red spoons	14	eight new books	9	eight green hats	17	eight new spoons
33	eight small hats	96	ten old toys	16	eight new shoes	27	eight red hats
48	nine green toys	100	ten red shoes	60	nine old toys	38	nine big books
50	nine new books	114	three big toys	64	nine red shoes	97	ten red bikes
58	nine old shoes	127	three old bikes	108	ten small toys	115	three green bikes
90	ten new toys	159	twelve new hats	137	three red spoons	129	three old hats
142	three small shoes	173	twelve red spoons	154	twelve green shoes	134	three red books
151	twelve green bikes	176	twelve small books	163	twelve old bikes	157	twelve new bikes
167	twelve old spoons	183	two big hats	181	two big bikes	178	twelve small shoes
182	two big books	191	two green spoons	197	two new spoons	190	two green shoes
				200	two old books		
List 9		**List 10**		**List 11**		**List 12**	
1	eight big bikes	3	eight big hats	12	eight green toys	11	eight green spoons
18	eight new toys	59	nine old spoons	25	eight red bikes	26	eight red books
51	nine new hats	66	nine red toys	53	nine new spoons	63	nine red hats
92	ten old books	76	ten big shoes	98	ten red books	78	ten big toys
102	ten red toys	120	three green toys	105	ten small hats	91	ten old bikes
130	three old shoes	125	three new spoons	143	three small spoons	131	three old spoons
135	three red hats	166	twelve old shoes	146	twelve big books	144	three small toys
152	twelve green books	175	twelve small bikes	160	twelve new shoes	148	twelve big shoes
179	twelve small spoons	193	two new bikes	168	twelve old toys	195	two new hats
187	two green bikes	206	two red books	199	two old bikes	205	two red bikes
List 13		**List 14**		**List 15**		**List 16**	
19	eight old bikes	20	eight old books	24	eight old toys	22	eight old shoes
30	eight red toys	36	eight small toys	37	nine big bikes	39	nine big hats
46	nine green shoes	45	nine green hats	57	nine old hats	62	nine red books
49	nine new bikes	86	ten new books	84	ten green toys	77	ten big spoons
74	ten big books	121	three new bikes	87	ten new hats	82	ten green shoes
99	ten red hats	136	three red shoes	116	three green books	133	three red bikes
141	three small hats	149	twelve big spoons	124	three new shoes	140	three small books
155	twelve green spoons	165	twelve old hats	169	twelve red bikes	153	twelve green hats
196	two new shoes	209	two red spoons	210	two red toys	162	twelve new toys
215	two small spoons	211	two small bikes	212	two small books	216	two small toys

APPENDIX B:

Hearing Thresholds

B.1 Average PTA thresholds of participants in this study.



Note. Error bars represent the standard deviation of the threshold at each frequency.

APPENDIX C:

Part A: Data Table

C.1 SRT and slope values for each of Jenkins-Foreman's (2018) UCAMST-P sentence lists for the open-set response format.

3x6, Open-set, Auditory-Alone (combined with/without practice conditions)																
SRT	**List 1**	**List 2**	**List 3**	**List 4**	**List 5**	**List 6**	**List 7**	**List 8**	**List 9**	**List 10**	**List 11**	**List 12**	**List 13**	**List 14**	**List 15**	**List 16**
Sentence 01	-13.6	-1.8	-11.4	-8.6	-11.3	-10.4	-6.3	-13.4	-11.6	-9.4	-8.7	-7.1	-8.9	-11.0	-9.4	-9.4
Sentence 02	-9.2	-12.6	-9.5	-14.9	-8.9	-9.3	-11.2	-11.6	-7.5	-10.4	-11.5	-10.2	-8.9	-9.3	-9.4	-9.3
Sentence 03	-11.4	-10.3	-10.6	-9.5	-11.0	-11.3	-9.0	-10.1	-7.8	-10.0	-9.1	-8.8	-7.6	-6.7	-8.3	-10.4
Sentence 04	-7.9	-4.2	-11.3	-11.7	-8.6	-11.4	-8.6		-8.2	-10.7	-11.2	-10.6	-7.5	-7.3	-7.6	-13.1
Sentence 05	-11.5	-11.6	-8.6	-9.7	-11.4	-11.0	-8.0	-11.6	-11.2	-6.0	-6.8	-18.6	-10.2	-9.3	-6.7	-8.3
Sentence 06	-11.5	-10.0	-6.8	-9.6	-7.3	-8.4	-11.3	-7.1	-11.3	-11.1	-11.4	-12.3	-9.6	-11.9	-8.6	-10.4
Sentence 07	-11.4	-9.0	-11.2	-8.8	-8.9		-11.1	-11.8	-10.1	-11.6	-7.4	-18.6	-9.8	-9.5	-8.9	-11.3
Sentence 08	-8.4	-8.7	-11.8	-11.0	-9.4	-6.5	-9.0	-16.9	-11.1	-6.6	-9.6	-8.2	-7.6	-8.8	-10.5	-7.7
Sentence 09	-11.3	-7.1	-9.0	-13.6	-11.3		-11.6	-8.7	-8.7	-4.5	-11.2	-4.9	-6.3	-11.2	-9.4	-11.1
Sentence 10	-8.8	-7.0	-11.4	-9.2	-11.0	-10.0	-9.5	-10.2	-10.8	-7.4	-11.6	-10.9	-8.6	-9.7	-8.6	-8.1
Sentence 11				-7.9			-11.5									
Mean	-10.5	-8.2	-10.2	-10.4	-9.9	-9.8	-9.7	-11.3	-9.9	-8.8	-9.8	-11.0	-8.5	-9.5	-8.8	-9.9
StDev	± 1.8	± 3.3	± 1.6	± 2.2	± 1.5	± 1.7	± 1.7	± 2.8	± 1.6	± 2.5	± 1.8	± 4.5	± 1.2	± 1.6	± 1.1	± 1.7
Slope	**List 1**	**List 2**	**List 3**	**List 4**	**List 5**	**List 6**	**List 7**	**List 8**	**List 9**	**List 10**	**List 11**	**List 12**	**List 13**	**List 14**	**List 15**	**List 16**
Sentence 01	8.6	0.6	99.8	256.6	102.5	14.8	32.5	4.7	6.1	5.1	22.3	15.6	24.3	5.1	20.5	10.2
Sentence 02	14.4	11.3	16.1	5.2	24.3	7.6	99.4	4.5	42.4	9.1	98.6	12.3	24.3	7.6	20.5	5.2
Sentence 03	99.8	8.4	16.4	4.9	107.8	12.3	15.4	11.4	18.7	6.9	18.9	14.3	43.2	13.5	17.1	14.8
Sentence 04	17.8	9.5	102.5	99.1	256.6	99.0	256.6		20.9	5.9	104.6	11.2	15.9	7.9	18.8	4.9
Sentence 05	102.6	104.0	256.6	9.2	99.0	5.1	16.8	4.5	103.5	162.2	35.5	2.5	12.3	7.6	34.9	10.9
Sentence 06	104.7	7.2	35.8	12.8	39.9	10.6	101.4	8.9	102.1	11.6	99.7	8.2	8.6	91.9	256.6	5.8
Sentence 07	99.0	12.0	104.6	9.0	17.5		106.8	101.0	15.9	10.6	41.3	2.5	13.7	4.9	13.5	6.6
Sentence 08	14.4	8.3	93.5	107.8	20.9	6.7	15.4	1.6	105.4	14.1	16.8	5.1	43.2	12.3	15.6	14.5
Sentence 09	102.5	15.6	9.6	8.6	102.5	0.0	11.7	9.4	24.3	5.6	104.6	10.6	32.7	8.9	20.5	13.0
Sentence 10	12.3	11.4	99.8	14.4	107.8	6.3	18.8	6.0	13.7	12.3	107.6	8.1	256.6	9.2	256.6	4.9
Sentence 11				9.2			107.9									
Mean	57.6	18.8	83.5	48.8	87.9	18.0	71.2	16.9	45.3	24.3	65.0	9.0	47.5	16.9	67.5	9.1
StDev	± 46.6	± 30.2	± 73.2	± 78.6	± 71.2	± 30.7	± 74.6	± 31.7	± 41.3	± 48.5	± 40.8	± 4.6	± 74.5	± 26.5	± 99.9	± 4.1

APPENDIX C:

Part A: Data Table

C.2 SRT and slope values for each of Jenkins-Foreman's (2018) UCAMST-P sentence lists for the closed-set response format.

3x6, Closed-set, Auditory-Alone (combined with/without practice conditions)																
SRT	**List 1**	**List 2**	**List 3**	**List 4**	**List 5**	**List 6**	**List 7**	**List 8**	**List 9**	**List 10**	**List 11**	**List 12**	**List 13**	**List 14**	**List 15**	**List 16**
Sentence 01	-14.1	-11.6	-13.8	-10.7	-14.0	-11.5	-8.4	-14.0	-13.6	-14.4	-11.3	-11.4	-12.6	-10.2	-13.6	-12.8
Sentence 02	-10.0	-12.8	-12.9	-13.9	-13.2	-12.0	-9.2	-14.0	-11.8	-12.6	-12.3	-12.3	-14.3	-11.5	-11.9	-11.6
Sentence 03	-12.7	-12.1	-13.8	-12.7	-10.8	-13.8	-13.7	-13.5	-9.6	-13.4	-13.6	-11.3	-9.2	-10.1	-11.9	-12.2
Sentence 04	-11.3	-10.2	-12.6	-13.5	-11.3	-10.0	-13.7	-10.1	-10.8	-13.4	-10.6	-8.9	-10.7	-11.7	-10.5	-12.6
Sentence 05	-11.8	-11.8	-9.9	-9.2	-13.7	-13.8	-11.5	-10.2	-13.9	-10.2	-10.3	-15.4	-13.8	-11.7	-10.4	-9.2
Sentence 06	-11.8	-13.9	-7.0	-11.0	-10.8	-10.4	-13.6	-11.6	-13.7	-5.8	-13.6	-14.2	-10.0	-13.9	-9.2	-13.7
Sentence 07	-10.8	-9.2	-11.8	-12.6	-13.0	-13.8	-12.0	-13.6	-13.9	-11.6	-11.6	-13.5	-11.7	-8.4	-10.4	-13.7
Sentence 08	-10.4	-12.6	-13.8	-11.5	-11.7	-10.9	-13.8	-11.9	-12.3	-12.2	-11.7	-11.0	-10.7	-12.1	-12.7	-10.7
Sentence 09	-13.9	-10.0	-12.9	-13.9	-12.2	-10.4	-12.5	-11.1	-11.3	-8.8	-13.8	-8.2	-8.3	-10.7	-11.4	-14.2
Sentence 10	-11.3	-11.2	-14.0	-13.5	-10.8	-10.1	-10.8	-10.2	-13.7	-11.3	-13.8	-10.4	-13.7	-10.1	-11.9	-13.1
Sentence 11				-10.5			-10.4									
Mean	-11.8	-11.5	-12.2	-12.1	-12.2	-11.7	-11.8	-12.0	-12.5	-11.4	-12.3	-11.7	-11.5	-11.0	-11.4	-12.4
StDev	± 1.4	± 1.4	± 2.2	± 1.6	± 1.2	± 1.6	± 1.9	± 1.6	± 1.5	± 2.6	± 1.4	± 2.3	± 2.1	± 1.5	± 1.3	± 1.5
Slope	**List 1**	**List 2**	**List 3**	**List 4**	**List 5**	**List 6**	**List 7**	**List 8**	**List 9**	**List 10**	**List 11**	**List 12**	**List 13**	**List 14**	**List 15**	**List 16**
Sentence 01	90.4	11.1	86.5	5.3	103.6	9.8	8.5	101.4	90.7	6.7	13.6	11.7	12.8	8.3	87.1	8.7
Sentence 02	12.1	8.7	8.5	87.1	8.0	8.8	9.9	91.8	11.9	12.8	14.4	13.6	9.6	7.1	16.8	11.1
Sentence 03	13.5	14.0	86.5	6.6	7.6	92.3	89.4	89.6	10.9	90.6	88.1	17.0	14.2	10.7	16.8	9.8
Sentence 04	18.1	9.2	12.8	89.6	9.0	10.4	89.4	19.1	15.2	5.8	22.7	11.4	10.7	10.5	7.4	12.8
Sentence 05	11.9	8.0	159.3	1.8	87.4	86.5	14.5	8.5	87.3	9.2	8.1	8.2	85.6	10.5	9.8	9.9
Sentence 06	16.2	91.6	8.8	9.5	10.1	9.0	87.1	17.1	89.4	2.2	89.1	9.7	6.6	91.9	9.9	87.6
Sentence 07	15.2	14.2	8.0	12.8	11.8	85.7	8.8	86.7	87.3	11.1	9.7	5.7	15.5	7.3	9.8	89.4
Sentence 08	13.4	12.8	86.5	16.3	10.9	7.9	86.5	15.7	10.6	2.1	16.4	9.8	14.4	9.5	14.1	7.8
Sentence 09	91.6	6.6	8.5	10.1	13.8	9.0	11.9	11.0	13.3	13.0	86.6	13.4	11.9	5.3	14.0	83.0
Sentence 10	13.3	12.5	103.6	89.6	10.1	6.3	8.1	8.5	89.4	6.5	85.8	8.4	89.4	13.6	12.5	11.7
Sentence 11				8.1			13.4									
Mean	29.6	18.9	56.9	30.6	27.2	32.6	38.9	44.9	50.6	16.0	43.4	10.9	27.1	17.5	19.8	33.2
StDev	± 32.4	± 25.7	± 54.4	± 37.5	± 36.2	± 38.4	± 39.1	± 41.1	± 40.3	± 26.5	± 38.0	± 3.3	± 32.0	± 26.2	± 23.8	± 37.0

APPENDIX D:

Adjusted Results from Lay (2019)

D.1 SRT and slope values for each of Lay's (2019) UCAMST-P sentence lists for the open-set response format.

3x6, Open-set, Auditory-Alone (combined with/without practice conditions) - Lay (2018)												
SRT	**List 1**	**List 2**	**List 3**	**List 4**	**List 5**	**List 6**	**List 7**	**List 8**	**List 9**	**List 10**	**List 11**	**List 12**
Sentence 01	-6.5	-9.0	-8.9	-7.3	-11.6	-11.5	-11.5	-10.3	-8.7	-8.3	-9.7	-11.3
Sentence 02	-9.4	-11.3	-11.0	-6.2	-8.7	-7.4	-7.9	-10.7	-8.8	-11.4	-8.4	-10.4
Sentence 03	-7.6	-9.5	-7.0	-9.6	-10.8	-9.6	-11.4	-9.2	-11.2	-6.6	-12.6	-10.4
Sentence 04	-10.5	-10.2	-11.3	-10.2	-7.2	-11.2	-8.8	-11.2	-9.7	-9.4	-11.6	-11.6
Sentence 05	-8.9	-10.9	-11.0	-7.3	-7.8	-11.4	-13.6	-11.7	-10.2	-11.1	-10.0	-13.6
Sentence 06	-10.2	-11.2	-11.4	-8.9	-11.3	-11.2	-11.4	-8.3	-11.3	-9.0	-14.9	-8.3
Sentence 07	-9.4	-9.0	-9.5	-9.8	-11.1	-8.7	-11.3	-9.3	-11.1	-11.4	-10.1	-7.9
Sentence 08	-9.4	-6.2	-10.6	-8.9	-8.2	-9.1	-11.5	-9.6	-11.4	-8.3	-7.7	-9.4
Sentence 09	-8.3	-8.0	-11.8	-10.2	-10.1	-7.1	-9.2	-11.0	-10.4	-11.3	-8.8	-10.6
Sentence 10	-8.3	-11.1	-11.3	-7.5	-11.2	-6.6	-8.4	-6.0	-6.7	-8.9	-9.0	-10.4
Mean	-8.9	-9.7	-10.4	-8.6	-9.8	-9.4	-10.5	-9.7	-10.0	-9.6	-10.3	-10.4
StDev	± 1.2	± 1.7	± 1.5	± 1.4	± 1.7	± 1.9	± 1.8	± 1.7	± 1.5	± 1.7	± 2.2	± 1.7
Slope	**List 1**	**List 2**	**List 3**	**List 4**	**List 5**	**List 6**	**List 7**	**List 8**	**List 9**	**List 10**	**List 11**	**List 12**
Sentence 01	45.5	15.4	17.5	53.9	6.1	104.6	100.6	8.4	8.3	239.7	9.2	6.6
Sentence 02	20.5	101.4	107.8	43.3	24.3	12.3	17.8	5.9	12.3	103.1	10.6	5.8
Sentence 03	18.8	18.8	50.5	8.6	13.7	16.8	99.8	14.4	8.9	46.4	11.3	14.8
Sentence 04	15.6	164.9	102.6	12.3	53.0	104.6	12.3	104.6	9.2	20.9	107.5	10.6
Sentence 05	13.5	8.1	107.8	53.9	18.7	101.9	8.6	99.1	12.3	13.0	7.2	8.6
Sentence 06	164.9	103.5	99.8	24.3	102.0	104.6	103.1	239.7	12.3	9.6	5.2	10.9
Sentence 07	20.5	15.4	16.1	13.7	107.6	22.3	102.6	7.6	11.6	99.8	11.4	9.2
Sentence 08	20.5	43.1	16.4	24.3	20.9	18.9	98.6	12.8	103.1	239.7	14.5	10.2
Sentence 09	239.7	16.8	95.5	164.9	15.9	51.9	14.4	107.8	14.8	102.6	14.3	11.2
Sentence 10	17.1	106.8	102.6	15.9	103.5	46.2	14.4	161.7	13.5	24.3	12.0	9.1
Mean	57.7	59.4	71.6	41.5	46.6	58.4	57.2	76.2	20.6	89.9	20.3	9.7
StDev	± 78.7	± 55.1	± 41.4	± 46.5	± 41.7	± 41.1	± 46.1	± 80.6	± 29.1	± 87.3	± 30.8	± 2.5

APPENDIX D:

Adjusted Results from Lay (2019)

D.2 SRT and slope values for each of Lay's (2019) UCAMST-P sentence lists for the closed-set response format.

3x6, Closed-set, Auditory-Alone (combined with/without practice conditions) - Lay (2018)												
SRT	**List 1**	**List 2**	**List 3**	**List 4**	**List 5**	**List 6**	**List 7**	**List 8**	**List 9**	**List 10**	**List 11**	**List 12**
Sentence 01	-10.4	-13.7	-13.0	-9.2	-13.6	-12.3	-11.8	-12.1	-12.6	-11.3	-9.2	-13.7
Sentence 02	-11.9	-13.6	-10.8	-8.3	-11.3	-11.3	-11.3	-13.4	-12.1	-13.7	-10.4	-13.7
Sentence 03	-10.5	-10.8	-10.8	-10.0	-13.7	-11.7	-12.7	-13.5	-10.7	-7.0	-12.8	-12.2
Sentence 04	-12.7	-13.7	-12.2	-13.8	-11.8	-10.6	-11.3	-11.8	-10.1	-11.7	-13.8	-11.6
Sentence 05	-10.4	-10.4	-10.8	-10.7	-9.6	-13.6	-14.1	-13.5	-12.3	-14.2	-13.9	-13.9
Sentence 06	-9.2	-9.2	-13.8	-14.3	-13.7	-13.8	-10.8	-10.7	-13.8	-12.9	-13.9	-9.2
Sentence 07	-13.6	-13.8	-12.9	-11.7	-12.3	-11.3	-13.9	-11.7	-5.8	-14.0	-13.5	-10.5
Sentence 08	-11.4	-8.4	-13.8	-12.6	-10.8	-13.6	-11.8	-11.0	-10.0	-9.9	-10.7	-12.8
Sentence 09	-11.9	-11.5	-13.8	-13.7	-13.9	-11.6	-10.0	-11.5	-11.5	-12.6	-11.3	-8.9
Sentence 10	-11.9	-12.0	-14.0	-10.7	-13.9	-10.3	-10.4	-10.2	-10.1	-13.2	-9.2	-12.6
Mean	-11.4	-11.7	-12.6	-11.5	-12.5	-12.0	-11.8	-11.9	-10.9	-12.1	-11.9	-11.9
StDev	± 1.3	± 2.0	± 1.3	± 2.1	± 1.5	± 1.3	± 1.4	± 1.2	± 2.2	± 2.2	± 1.9	± 1.8
Slope	**List 1**	**List 2**	**List 3**	**List 4**	**List 5**	**List 6**	**List 7**	**List 8**	**List 9**	**List 10**	**List 11**	**List 12**
Sentence 01	9.8	89.4	11.8	14.2	90.7	14.4	16.2	14.0	12.8	9.0	1.8	89.4
Sentence 02	16.8	87.1	7.6	11.9	13.3	6.5	18.1	5.8	9.5	87.4	9.0	87.6
Sentence 03	7.4	8.1	10.1	6.6	89.4	16.4	13.5	89.6	5.3	8.8	8.7	9.8
Sentence 04	14.1	89.4	13.8	85.6	11.9	22.7	13.3	8.0	13.6	10.9	85.8	11.1
Sentence 05	9.8	8.4	10.1	14.4	10.9	89.1	90.4	89.6	13.6	83.0	91.6	10.1
Sentence 06	9.9	9.9	86.5	9.6	89.4	86.6	15.2	5.3	92.3	8.5	87.1	9.9
Sentence 07	87.1	86.5	8.5	15.5	10.6	13.6	91.6	10.5	2.2	103.6	89.6	8.1
Sentence 08	14.0	8.5	86.5	12.8	15.2	88.1	11.9	9.5	10.4	159.3	7.8	8.7
Sentence 09	12.5	14.5	86.5	89.4	87.3	9.7	12.1	16.3	9.8	12.8	17.0	11.4
Sentence 10	16.8	8.8	103.6	10.7	87.3	8.1	13.4	9.2	10.7	8.0	14.2	12.8
Mean	19.8	41.1	42.5	27.1	50.6	35.5	29.6	25.8	18.0	49.1	41.3	25.9
StDev	± 23.8	± 40.5	± 41.9	± 32.0	± 40.3	± 36.4	± 32.4	± 33.8	± 26.3	± 54.9	± 40.9	± 33.0

D.3 Results of the univariate ANOVA for list equivalence (SRT).

Condition	Sum of squares	df	Mean square	*F*	*p*	η_p^2
AA, Open	40.281	11	3.662	1.310	.229	.118
AA, Closed	22.486	11	2.044	.686	.749	.065

Note. df = degrees of freedom; F = F-ratio; p = p-value; η_p^2 *= partial eta-squared.*

D.4 Results of the Kruskal-Wallis H test for list equivalence (slope).

Condition	df	χ^2	*p*
AA, Open	11	39.328	<.001
AA, Closed	11	9.450	.580

Note. df = degrees of freedom

D.5 Results of the univariate ANOVA for condition equivalence (SRT).

Condition	**Sum of squares**	**df**	**Mean square**	***F***	***p***	η_p^2
AA	259.277	1	259.277	89.812	<.001	.294

Note. ANOVA = Analysis of Variance; AA = auditory-alone; df = degrees of freedom; F = F-ratio; p = p-value; η_p^2 *= partial eta-squared.*

D.6 Results of the Mann-Whitney test for condition equivalence (slope).

Condition	***U***	***p***
AA	5,088	<.001

Note. U = Mann-Whitney; *p = Asymp Sig.*

APPENDIX E:

Part B: Data Tables

E.1 Average time per button press (s) between words for when the three-word and five-word sentences are presented in the AA, closed-set modality, with practice.

Time per button press (s) - Closed set, auditory-alone, with practice

Time per button press (s)	Low SNR (harder)					Higher SNR (easier)				
	1st word	2nd word	3rd word			1st word	2nd word	3rd word		
3x6	**1.44**	**0.63**	**0.66**			**1.19**	**0.73**	**0.91**		
	± 0.49	**± 0.39**	**± 0.45**			**± 0.37**	**± 0.36**	**± 0.65**		
	1st word	2nd word	3rd word	4th word	5th word	1st word	2nd word	3rd word	4th word	5th word
5x10	**1.77**	**0.75**	**0.71**	**0.74**	**0.79**	**1.62**	**1.07**	**0.94**	**1.30**	**1.40**
	± 0.64	**± 0.57**	**± 0.44**	**± 0.48**	**± 0.64**	**± 0.50**	**± 0.47**	**± 0.42**	**± 0.65**	**± 0.91**

E.2 Average time per button press (s) between words for when the three-word and five-word sentences are presented in the AA, closed-set modality, without practice.

Time per button press (s) - Closed set, auditory-alone, no practice

Time per button press (s)	Low SNR (harder)					Higher SNR (easier)				
	1st word	2nd word	3rd word			1st word	2nd word	3rd word		
3x6	**1.61**	**0.66**	**0.75**			**1.21**	**0.77**	**0.85**		
	± 0.66	**± 0.44**	**± 0.57**			**± 0.32**	**± 0.47**	**± 0.50**		
	1st word	2nd word	3rd word	4th word	5th word	1st word	2nd word	3rd word	4th word	5th word
5x10	**1.69**	**0.73**	**0.77**	**0.80**	**0.85**	**1.74**	**1.21**	**1.01**	**1.33**	**1.40**
	± 0.78	**± 0.44**	**± 0.57**	**± 0.48**	**± 0.68**	**± 0.64**	**± 0.87**	**± 0.61**	**± 0.69**	**± 0.84**